



USAID
FROM THE AMERICAN PEOPLE



TIME TO LEARN ENDLINE EVALUATION REPORT

February 2017

This publication was produced for the United States Agency for International Development Time to Learn Project. It was prepared by Zachariah Falconer-Stout, Rebecca Frischkorn, and Lynne Miller Franco, Time to Learn, EnCompass LLC.

Prepared for the United States Agency for International Development
Contract No. AID-611-C-12-00002

February 15, 2017

Authored by:

EnCompass LLC
1451 Rockville Pike, Suite 600
Rockville, MD 20852

Phone: +1 301-287-8700

Fax: +1 301-685-3720

www.encompassworld.com

TIME TO LEARN ENDLINE EVALUATION REPORT

DISCLAIMER

This evaluation was made possible by the support of the American people through the United States Agency for International Development (USAID). It was produced by Encompass LLC for the Time to Learn project, which was funded by USAID in Zambia under Contract No. AID-611-C-12-00002. The contents of this evaluation are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

Time to Learn Project

Time to Learn (TTL) was funded by USAID in Zambia under Contract No. AID 611-C-12-00002, awarded on March 1, 2012. TTL was implemented by Education Development Center, Inc. (EDC), in collaboration with the Campaign for Female Education (Camfed), the Forum for African Women Educationalists in Zambia (FAWEZA), and EnCompass LLC. The project assisted the Zambian Ministry of General Education (MOGE) through a 5-year national program to provide an equitable standard of education service for vulnerable learners, improve reading skills, and implement practical strategies to strengthen school quality and promote community engagement in community schools.

CONTENTS

ACRONYMS	vii
EVALUATION TEAM	viii
EXECUTIVE SUMMARY	x
Evaluation Purpose and Questions	x
Project Background	x
Evaluation Design, Methods, and Limitations	x
Findings, Conclusions, and Recommendations.....	xi
1. EVALUATION PURPOSE AND QUESTIONS.....	1
1.1. Evaluation Purpose	1
1.2. Evaluation Questions.....	2
2. PROJECT BACKGROUND	3
2.1. Project Context	3
2.2. TTL Interventions.....	4
3. EVALUATION DESIGN, METHODS, AND LIMITATIONS.....	11
3.1. Evaluation Design.....	11
3.2. Data Collection Methods and Data Collectors.....	12
3.3. Analysis	14
3.4. Limitations.....	15
4. DESCRIPTION OF THE ENDLINE SAMPLE.....	16
4.1. Characteristics of the Schools in the Sample.....	16
4.2. Characteristics of Learners in EGRA Sample.....	20
4.3. Characteristics of Teachers Participating in Literacy Lesson Classroom Observation.....	22
4.4. Characteristics of Community School Head Teacher Respondents.....	23
5. FINDINGS	25
5.1. Reading Achievement among Community School Learners.....	26
5.2. MOGE Support to Community Schools	39
5.3. Literacy Instruction in Community Schools.....	46
5.4. Factors Associated with Learner Outcomes	54
6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	57
6.1. Summary Responses to Evaluation Questions.....	57
6.2. Conclusions	59
6.3. Recommendations	62
ANNEX 1. DETAILED EGRA RESULTS.....	64
ANNEX 2. CLASSROOM OBSERVATION SUMMARY TABLES	92
ANNEX 3. DETAILED EVALUATION METHODS.....	97
ANNEX 4. EQUATING OF EGRA PASSAGES.....	108

ANNEX 5.	EVALUATION QUESTIONS BY DATA SOURCE AND INDICATOR	140
ANNEX 6.	REFERENCES	142
ANNEX 7.	DATA COLLECTION TOOLS	144

ACRONYMS

Camfed	Campaign for Female Education	MOGE	Ministry of General Education
CASAS	Center for Advanced Studies of African Society	OGCS	<i>Operational Guidelines for Community Schools</i>
CPD	Continuing professional development	PCSC	Parent community school committee
ECZ	Examinations Council of Zambia	SACMEQ	Southern and Eastern African Consortium for Monitoring Educational Quality
EGRA	Early grade reading assessment	TTL	Time to Learn
FAWEZA	Forum for African Women Educationalists in Zambia	USAID	United States Agency for International Development
IR	Intermediate result	ZIC	Zonal in-service coordinator

EVALUATION TEAM

ENCOMPASS CORE TEAM

Zachariah Falconer-Stout, Technical Lead

Rebecca Frischkorn, Team Lead

Lynne Miller Franco, Senior Research and Evaluation Advisor

Abby Ladd, Senior Evaluation Specialist

Han Bao, Psychometrician and Equating Statistician

Armando Levinson, Sampling Statistician

Pragati Godbole, Data Analyst

Maria Aguirre, Data Analyst

ENDLINE DATA COLLECTORS

TEAM LEADS

Godfrey Chitalu, Provincial Outreach Coordinator, TTL

Jane Lisimba, Provincial Outreach Coordinator, TTL

Paul Machona, Educational Leadership and Management Specialist, TTL

Kennedy Makulika, Monitoring and Evaluation Specialist, TTL

Janet Monde, District Resource Center Coordinator, MOGE

Felistas Moono, Provincial Outreach Coordinator, TTL

Michael Musonda, Provincial Outreach Coordinator, TTL

Annie Mwalusaka, District Resource Center Coordinator, MOGE

Richard Ndazye, District Resource Center Coordinator, MOGE

Valentine Yumba, Senior Education Standards Officer – Open and Distance Learning, MOGE

EGRA ASSESSORS

Anthony Chomba, District Statistician, MOGE

Paul Daka, Provincial Outreach Coordinator, TTL

Stanley Handema, Senior Education Standards Officer – Open and Distance Learning, MOGE

Kalimukwa Kalimukwa, Monitoring and Evaluation Assistant, TTL

Maureen Kasakula, Graduate Student, University of Zambia

John Mahachi, District Resource Center Coordinator, MOGE

Titus Miti, Provincial Statistician, MOGE

Rhodea Mudenda, District Resource Center Coordinator, MOGE

Bridget Mukwiza, District Resource Center Coordinator, MOGE

Madrine Musonda, District Resource Center Coordinator, MOGE

Wesley Mweemba, District Resource Center Coordinator, MOGE

Madelene Mwila, Graduate Student, University of Zambia

Mbololwe Nalumino, Assistant District Resource Center Coordinator, MOGE

Lucy Njobvu, Education Standards Officer –
General Inspection, MOGE

Francis Phiri, Provincial Outreach Coordinator,
TTL

Mampi Sakala, Monitoring and Evaluation
Assistant, TTL

Regina Siamusiye, Senior Education Standards
Officer – Languages, MOGE

Amos Simusokwe, Intern, TTL

Joshua Zulu, Graduate Student, University of
Zambia

CLASSROOM OBSERVERS AND HEAD TEACHER QUESTIONNAIRE ADMINISTRATORS

Mabel Banda, Monitoring and Evaluation
Assistant, TTL

Mumbi Chanda, Monitoring and Evaluation
Assistant, TTL

Martin Chibulika, Head Teacher, MOGE

Sheila Chiile, District Resource Center
Coordinator, MOGE

Chanda Chileshe, District Resource Center
Coordinator, MOGE

Kenneth Jinaina, Curriculum Development
Specialist – Research and Evaluation, MOGE

Titus Kalumba, Assistant District Resource
Center Coordinator, MOGE

Kebby Mabili, District Statistician, MOGE

Maurice Mulopo, Senior Education Officer –
Open and Distance Learning, MOGE

Maipambe Mutale, Intern, TTL

Sunday Simutowe, District Resource Center
Coordinator, MOGE

EXECUTIVE SUMMARY

EVALUATION PURPOSE AND QUESTIONS

The USAID Time to Learn (TTL) project's endline evaluation seeks to assess the degree to which TTL achieved its intended impact and outcomes and to draw conclusions about how the project's experiences can provide important learning for future education interventions in community schools in Zambia.

Measurement of TTL impacts and outcomes began in project year 1 (baseline in 2012, prior to project interventions), followed by a midline in year 3 (2014) and an endline in year 5 (2016, 4 years after project interventions began). These measurements included literacy levels among learners and other intermediate outcomes to determine if teachers, head teachers, and the Ministry of General Education (MOGE) were using TTL-promoted techniques, tools, and approaches.

This evaluation report focuses on three key evaluation questions, listed in [Exhibit 1](#), and uses data from all three measurement periods.

PROJECT BACKGROUND

TTL was a 5-year project that endeavored to improve reading among 500,000 grade 1–4 community school learners in 6 of Zambia's 10 provinces. TTL served more than 2,000 schools, covering three languages of instruction: ChiTonga, CiNyanja, and iCiBemba. TTL contributed to Goal 1 of the [USAID Education Strategy](#), "Improved reading skills for 100 million children in primary grades by 2017." The project aimed to create a favorable environment for effective implementation of the MOGE's policy to integrate community schools into the formal education system, while engaging a range of MOGE actors in ways that would facilitate sustaining and generalizing TTL interventions for scale-up.

TTL's development hypothesis held that community school learners would improve their reading performance as the result of (1) increased MOGE support to community schools; (2) improved literacy instruction and educational management at community schools; (3) improved parent community school committee (PCSC) governance, resource mobilization, and advocacy for quality reading instruction and support; (4) increased support for literacy in the home; and (5) increased community school access to age-appropriate, familiar-language teaching and learning materials. TTL sought to achieve these results through four intervention areas: MOGE capacity building, PCSC capacity building, teacher training, and development and dissemination of teaching and learning materials.

EVALUATION DESIGN, METHODS, AND LIMITATIONS

The evaluation team developed the 5-year evaluation design in consultation with the MOGE, the Examinations Council of Zambia (ECZ), USAID, and TTL project specialists. In keeping with USAID-recommended approaches for measuring reading gains, this evaluation uses an independently pooled,

repeated cross-sectional design. Responses to key evaluation questions 1 and 2 are based on comparisons with baseline and midline data. Question 3 is based on comparison to midline only, due to changes in the instrument after baseline.

School-based data collection for all three phases occurred at the same point in the school year, and included early grade reading assessment (EGRA), classroom observation (one literacy lesson per school), a head teacher questionnaire, and a self-administered questionnaire for MOGE District Education Board Secretaries.

Sampling was based on a two-stage sampling design, with target sample sizes set at provincial level and separate samples drawn for each province to enable estimation of reading results at that level. The number of learners sampled for EGRA was 1,308 at baseline, 1,683 at midline, and 1,798 at endline. The EGRA sample represented 73 community schools at baseline, 102 at midline, and 105 at endline. This sample of community schools, plus an additional 121 schools at baseline, provided classroom observation and head teacher interview data. Analysis involved descriptive statistics and bivariate and inferential techniques, and examined similarities and differences between males and females.

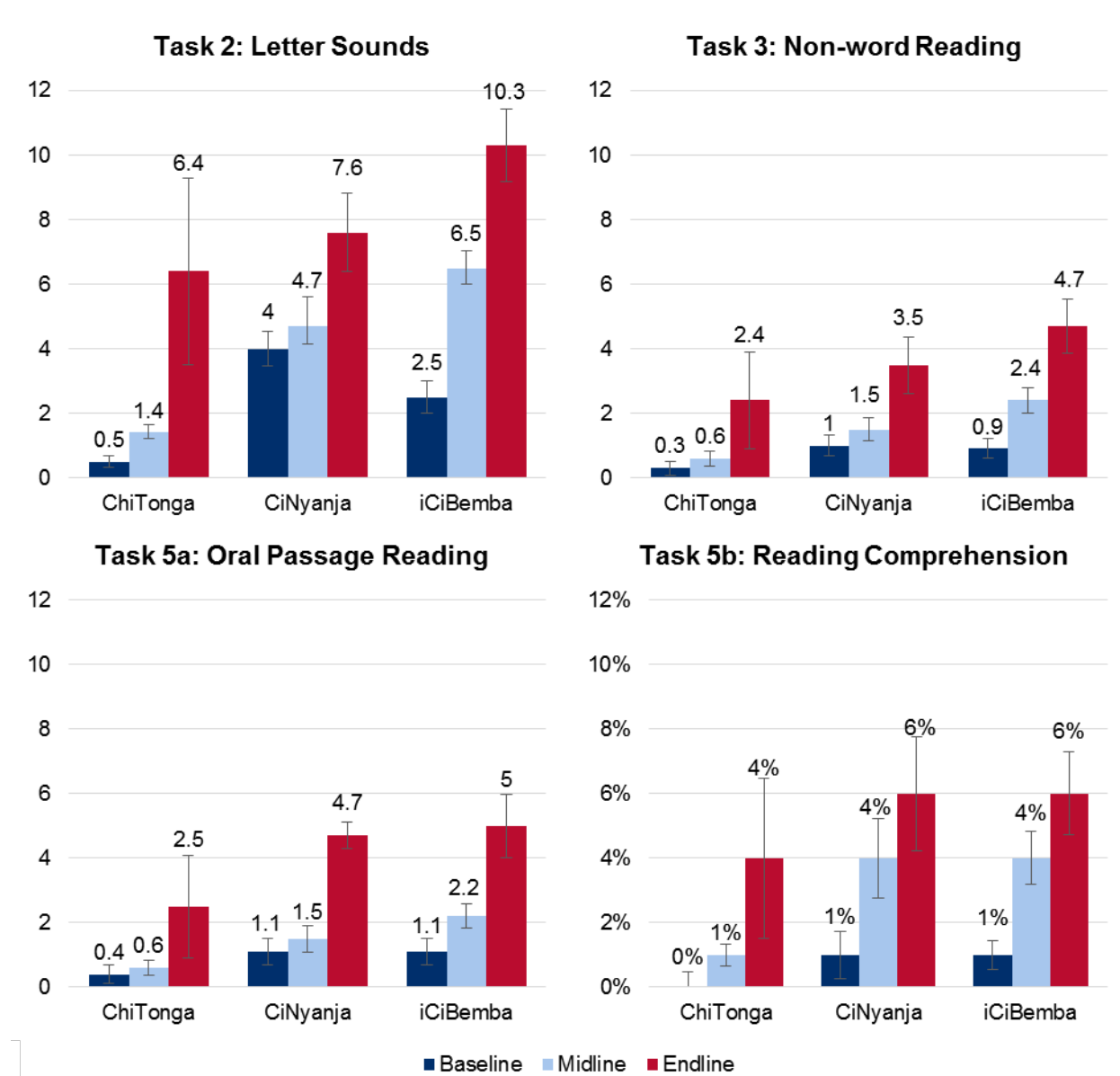
This evaluation is limited by the absence of a counterfactual (a non-intervention comparison or control group). The evaluation relies on contextual data to build evidence for TTL's contributions to improvements, but it cannot statistically *attribute* results to TTL alone. Additionally, the EGRA instruments do not allow comparison across languages of instruction and data for some schools, and classroom indicators are not available across all three phases.

FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

FINDINGS

A profound shift has occurred in the past 5 years in TTL community schools, where the majority of learners are now taking their first steps toward learning to read. Results among grade 2 learners in community schools have shown statistically significant increases in reading ability since baseline. Baseline values for EGRA were very low; the majority of learners could not sound a single letter, and 9 out of 10 learners could not read a single word. There has been growth in the proportion of learners who read at grade level, as defined by MOGE benchmarks, but the proportion is still low overall. Over the same period, the proportion of learners possessing emergent literacy skills has seen much larger increases (“emergent” is the MOGE benchmark immediately below “reader”). Many of the foundational literacy skills measured by EGRA have also shown large improvements, as measured by effect sizes of the change from baseline to endline. Improvements are particularly apparent for the fundamental reading competencies of sounding letters (EGRA task 2), decoding non-words (EGRA task 3), and reading fluency (EGRA task 5a), with more limited improvements for reading comprehension (EGRA task 5b), as [Exhibit 2](#) illustrates. The overall positive shifts are due to both a decrease in non-readers and an increase in the proportion of learners achieving at the higher level of “emergent readers.” However, no changes have been seen in listening comprehension since midline, and weak oral language proficiency may be impeding development of other literacy skills.

Exhibit 2: EGRA Mean Scores at Baseline, Midline, and Endline



Overall, gains in literacy skills are similar for male and female learners, indicating that parity in performance has been maintained as reading levels have improved, even as progress has been more varied across provinces and individual schools. Central and Copperbelt provinces have shown improvement across all EGRA tasks at endline, while the other provinces have had more mixed results across tasks and time. Additionally, variation in reading outcomes fluctuated from one school to the next at midline and endline; at some schools, almost all learners scored at or near zero, while other schools had comparatively exceptional learning outcomes among the learners sampled.

The MOGE has significantly increased the proportion of community schools it provides with teaching and learning materials, free basic materials, teacher training, monitoring, deployment of government teachers, and infrastructure. Although the proportion of schools receiving direct financial support from the MOGE has not increased since the baseline, the quality of monitoring support MOGE staff provides appears to be improving, with more visits, including classroom observation with feedback. At endline,

the MOGE was reported as the primary provider of support among sampled schools. Unsurprisingly, community school head teachers still reported that their schools need additional resources.

Classroom practices are shifting to strengthened literacy teaching methods and better balancing of lessons across the five literacy competencies. There have been significant observable changes from midline, with more observed lessons at endline, including reading comprehension and fluency activities. However, an imbalance remains between activities focused on basic phonics and those focused on developing comprehension skills, both oral and written. Most teachers reported using TTL-supported practices and MOGE materials and showed greater understanding of the MOGE policy on language of instruction, aligning to USAID-promoted practices and based on strong global evidence. Reported practices indicate greater use of TTL- and MOGE-promoted instructional approaches and classroom and school management practices.

The extent to which learners practice reading at home and at school, both alone and with adults, was a positive significant predictor of reading fluency and foundational skills in phonics and decoding; stronger oral language ability was also associated with higher reading outcomes. At the school level, rural schools and those with lower learner-to-teacher ratios were associated with improved learner reading levels. Notably, sex and socioeconomic status were not significant predictors of learners' reading ability in regression analysis. The improvement in reading scores over time remained significant even when controlling for these other factors, providing further evidence that ability is truly increasing.

CONCLUSIONS

Early grade reading acquisition is a complex process requiring multiple linguistic and cognitive skills. Research and evaluation conducted over the course of TTL make it clear, as does global evidence, that there is not only complexity in literacy acquisition, regardless of the context, but also that Zambia presents specific challenges to achieving grade-level reading performance, especially in its community schools. Before the MOGE's policy shift in teaching literacy (as implemented through TTL in community schools), reading scores in Zambia and much of southern Africa had been declining. Community schools, though they increase Zambian children's access to primary education, face additional challenges related to quality. Most community schools remain severely under-resourced in terms of basic educational inputs, even if they benefit from community commitment and volunteer teachers. Of the 105 randomly selected schools in the endline sample, the average teacher-to-learner ratio is 1:60, and more than half of community school teachers have no professional qualifications. **In this context, TTL, with USAID and the MOGE, has been working to improve reading and to reverse a previous negative trend. The extent of improvement observed in the TTL intervention area over 4 years demonstrates not only strong progress, but also remarkable progress.**

TTL has been an important actor in the development and implementation of policy changes regarding early grade literacy in Zambia. The project fostered the inclusion of the community school perspective and ensured that the MOGE incorporate community schools in sector-wide changes to educational practice. **The success of this strategy is seen in the strong growth in the MOGE's support to community schools over the course of TTL.**

Teachers increasingly used TTL-supported instructional practices, mirroring the changes in learner performance: Stronger learner results are observed in the areas teachers practice most during lessons (phonics and decoding), with less progress in those they practice less (oral language development and reading comprehension). The changes evident in teaching practices at endline reflect a continuation of the steady pace of change observed in the year 2 (2013) and year 4 (2015) performance evaluations and the midline (2014). In both performance evaluations, teachers stated that they were making the changes as a result of TTL training, and the midline and endline evaluations confirmed that these changes were

broad-based. **The emergence of this theme at each evaluation phase, confirmed by qualitative and quantitative data sources, provides strong evidence supporting TTL’s positive role.**

Learning performance shows endline improvements over baseline and midline, but based on the current rate of change (2012–2016), community schools are not improving reading outcomes at a pace sufficient to meet the MOGE’s 5-year targets (to be achieved by 2019) for the “reader” and “emergent reader” categories. In addition, the substantial variation in reading outcomes within and between schools reveals that future community school policy and support cannot take a one-size-fits-all approach to increasing performance at *all* schools and meet targets. More research is needed to understand how to leverage factors influencing performance at both learner and school levels. Moreover, weakness in listening comprehension in the language of instruction could be hindering progress in acquisition of reading fluency and reading comprehension.

In summary, **TTL interventions have shown that it is possible to improve support to community schools and classroom practice, and ultimately to improve reading scores among community school learners.** To maintain and expand emerging progress, this work needs to continue.

Regression analysis indicates intervention areas with promising potential to positively influence reading ability. Key factors associated with learner performance that could be targeted now include a greater emphasis on guided and individual reading practice, both in and out of school. More study is needed to better understand these factors and how interventions can effectively influence them.

RECOMMENDATIONS

1. The MOGE and its partners should target support to community schools in ways that recognize the wide variation within and among community schools.
2. USAID and others should support additional research to explore factors associated with weak oral language proficiency and ways it may be impeding development of other literacy skills among community school learners.
3. The MOGE should continue its integration of community schools into the Zambian education system and provide consistent distribution of resources, in line with the level of material need and appropriate level of support, to ensure educational quality across the diverse landscape of community schools.
4. The MOGE and its partners should provide additional guidance to teachers on structuring literacy lessons and integrating higher- and lower-order literacy skills into teaching practice in a way that produces the desired reading levels.

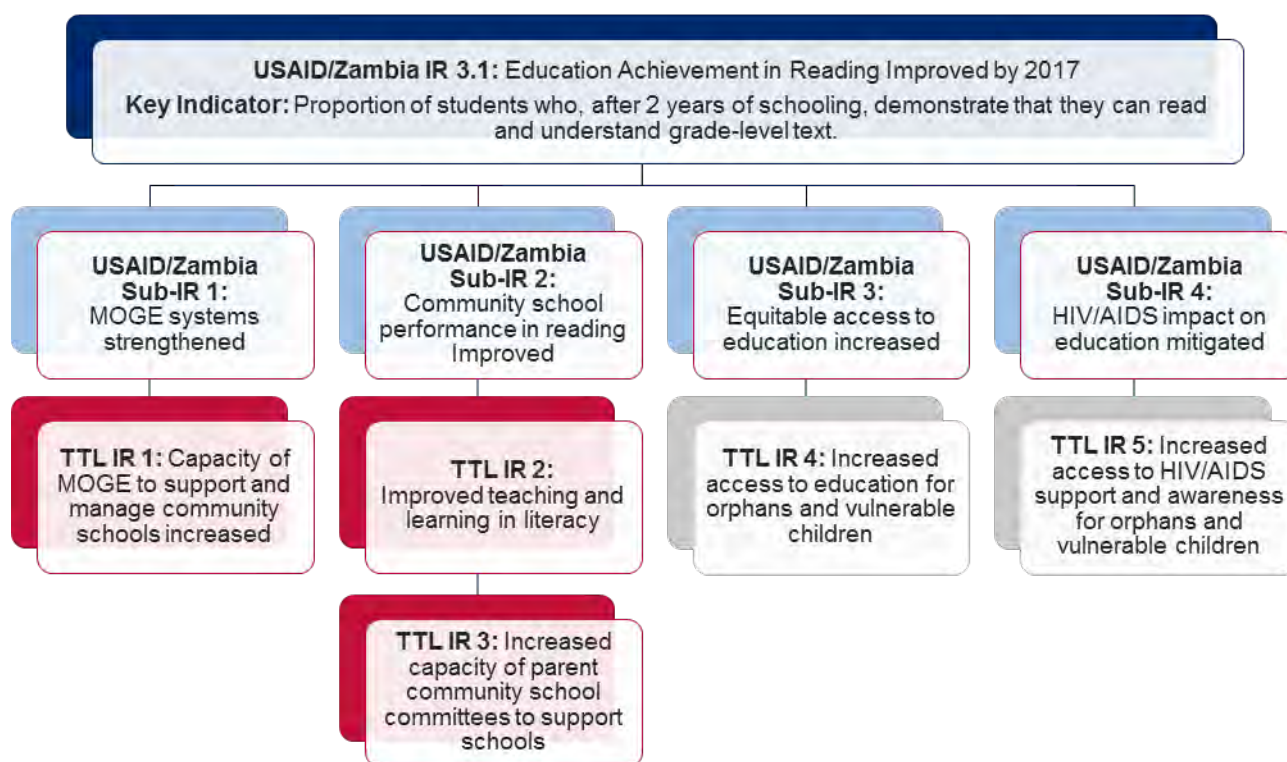
I. EVALUATION PURPOSE AND QUESTIONS

I.1. EVALUATION PURPOSE

This internal endline evaluation (conducted 4 years after interventions began) assesses the degree to which the USAID TTL project achieved its intended impact and outcomes and what lessons can be applied to future education interventions in Zambian community schools.

TTL conducted internal evaluations of reading and classroom practices and performance evaluations over the life of the project, for itself and its key stakeholders, to understand progress toward three of the five TTL intermediate results (IRs), as indicated in the red boxes in [Exhibit 3](#).

Exhibit 3: TTL Results Framework in the Context of USAID's Results Framework



Combined, these evaluations reflect a multilevel and sequential mixed-method approach so that TTL can assess its interventions at different points and from different perspectives, and understand project results holistically over time. This approach is illustrated in [Exhibit 3](#), in which the blue boxes indicate the three phases of the reading and classroom practice evaluation and the red boxes indicate performance evaluations.

Performance evaluations used qualitative and limited quantitative data to assess how TTL interventions were being implemented and perceived by key stakeholders. Conducted in project years 2 and 4, between the reading and classroom practice evaluation phases, the performance evaluations helped TTL understand *why* it was effective and *how* it could be improved.

The three reading and classroom practice evaluation phases were conducted in project year 1 (baseline), year 3 (midline), and year 5 (endline) to measure literacy levels among learners in primary grades (see [Exhibit 3](#), USAID/Zambia's IR 3.1 and other intermediate outcomes to determine if teachers, head teachers, and the MOGE were using TTL-promoted techniques. The reading and classroom practice evaluation activities used primarily quantitative data.

Exhibit 4: TTL Evaluation Timeline



I.2. EVALUATION QUESTIONS

The 2012 TTL evaluation design, developed before baseline data collection, conceptualizes five evaluation questions seeking to understand TTL's impact. The midline evaluation team refined these to three key questions and corresponding sub-questions, as listed in [Exhibit 5](#). These modifications were made with the TTL project team (documented in the midline evaluation implementation plan) to align the evaluation questions with project implementation since baseline, TTL's information needs, and USAID's priorities.

Exhibit 5: Evaluation Questions

Key Evaluation Question	Sub-questions
1. To what extent have TTL interventions improved early grade literacy achievement among male and female learners in community schools across six provinces compared to baseline? (TTL IR 2)	a) To what extent, and for how many learners, have TTL interventions increased reading skills among boys and girls across six provinces? (TTL indicator 2.1.1) b) What proportion of male and female learners across six provinces, in TTL-supported community schools, can read and understand the meaning of grade-level text after 2 years of primary schooling? (TTL indicator 2.1)
2. How has the MOGE's engagement with community schools changed since baseline? (TTL IR 1)	a) How many community schools receive more support from the MOGE compared to baseline? (TTL indicator 1.3) b) To what extent has the MOGE improved its monitoring of community schools compared to baseline? (TTL indicator 1.2)
3. To what extent are male and female teachers implementing TTL-supported literacy teaching methods? (TTL IR 2)	

2. PROJECT BACKGROUND

The 5-year TTL project endeavored to improve reading among 500,000 grade 1–4 community school learners in six of Zambia’s 10 provinces, as illustrated in [Exhibit 6](#) and [Exhibit 7](#). TTL contributes to Goal 1 of the [USAID Education Strategy](#), “Improved reading skills for 100 million children in primary grades by 2017,” which establishes early grade reading ability as a key determinant of retention and success in future grades.

TTL aimed to create a favorable environment for effective implementation of the MOGE policy of integrating community schools into the formal education system and to provide a wide range of MOGE actors with an opportunity to understand how to sustain and generalize TTL interventions for scale-up.

2.1. PROJECT CONTEXT

In 2002, Zambia declared free primary education by officially abolishing school fees for grades 1–7. According to the 2015 MOGE Educational Statistical Bulletin, male and female enrollment in early grades has increased steadily since then, a change that is largely attributed to community schools (MOGE 2016a: 24).

As their name implies, community schools are created and managed by communities through the PCSCs, which have primary responsibility for managing and supporting the schools. The majority of community schools serve grades 1–7. In 2011, the Zambia Education Act officially recognized community schools as one of four types of primary and secondary educational institutions. The MOGE subsequently launched a series of policies to integrate community schools into the formal education sector.

Today, community schools still serve Zambia’s most vulnerable rural and urban communities by helping to relieve overcrowding and by providing access to education in the most rural areas, where government schools are too distant for young learners to access. Yet accurate numbers of community schools are difficult to obtain. The MOGE counts some 2,400 registered community schools nationally, about 30 percent of primary schools across Zambia (MOGE 2016a). However, TTL’s effort to accurately document community schools in its catchment area found that this number underrepresented the presence of community schools. In 2016 alone, TTL served 2,279 active community schools with more than 400,000 enrolled learners in the six provinces where it worked, compared with 1,642 reported by the most recent annual school census data (MOGE 2016a). Including the additional TTL-supported community schools would increase the proportion of these schools to 44 percent of primary schools in those six TTL provinces.

Exhibit 6: Map of Provinces Where TTL Worked

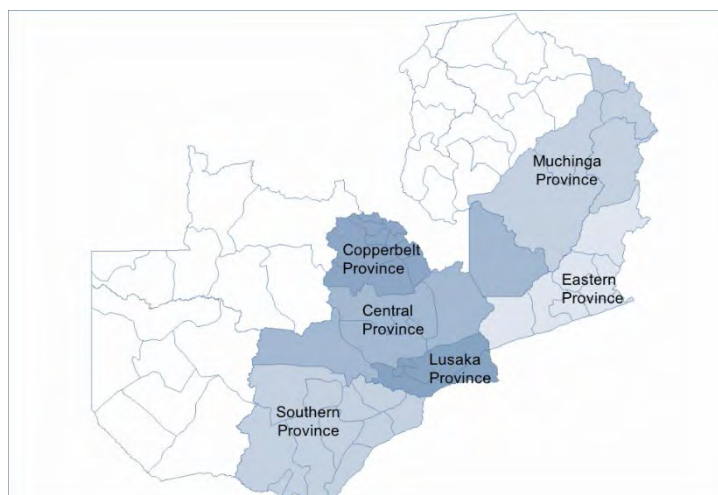


Exhibit 7: TTL Catchment Area Population

Province	Districts	Zones	Community	Teachers*	Learners*
----------	-----------	-------	-----------	-----------	-----------

Schools

Central	11	108	405	1,854	76,490
Copperbelt	10	101	381	2,112	86,587
Eastern	9	119	393	1,560	65,255
Lusaka	8	36	471	2,485	96,915
Muchinga	7	83	202	563	30,539
Southern	13	137	427	2,239	82,088
TOTAL	58	584	2,279	10,813	437,874

* Data for the number of teachers and learners are available for only 2,001 schools. (Source: 2016 TTL Monitoring Data)

Most community school teachers are locally recruited volunteers who receive small stipends from the community, a practice that generates local buy-in and makes these schools a low-cost alternative to government schools. Although these volunteer teachers are more likely to lack formal teacher training and generally have no more than a secondary school education, they often demonstrate high levels of commitment to their communities and accountability to the PCSC. This accountability is widely regarded as a major strength of Zambia's community schools (Frischkorn and Falconer-Stout 2016). The MOGE has committed to deploying trained government teachers to community schools and to monitoring these schools. Registered community schools are eligible to receive government assistance in the form of continuing professional development (CPD), such as in-service training; grants, teaching and learning materials; free basic materials, such as chalk, learner books, and pencils; and infrastructure.

While increasing community school enrollment has helped Zambia achieve the Education for All universal primary enrollment targets, educational quality has remained a challenge across the primary education sector. Zambian primary school learning outcomes have remained among the lowest in Sub-Saharan Africa, and low levels of literacy attainment are a particular challenge. The 2014 Grade 5 National Assessment Survey found that only one-third of Zambia's learners met national literacy standards by the end of grade 5 (MOGE 2015b).

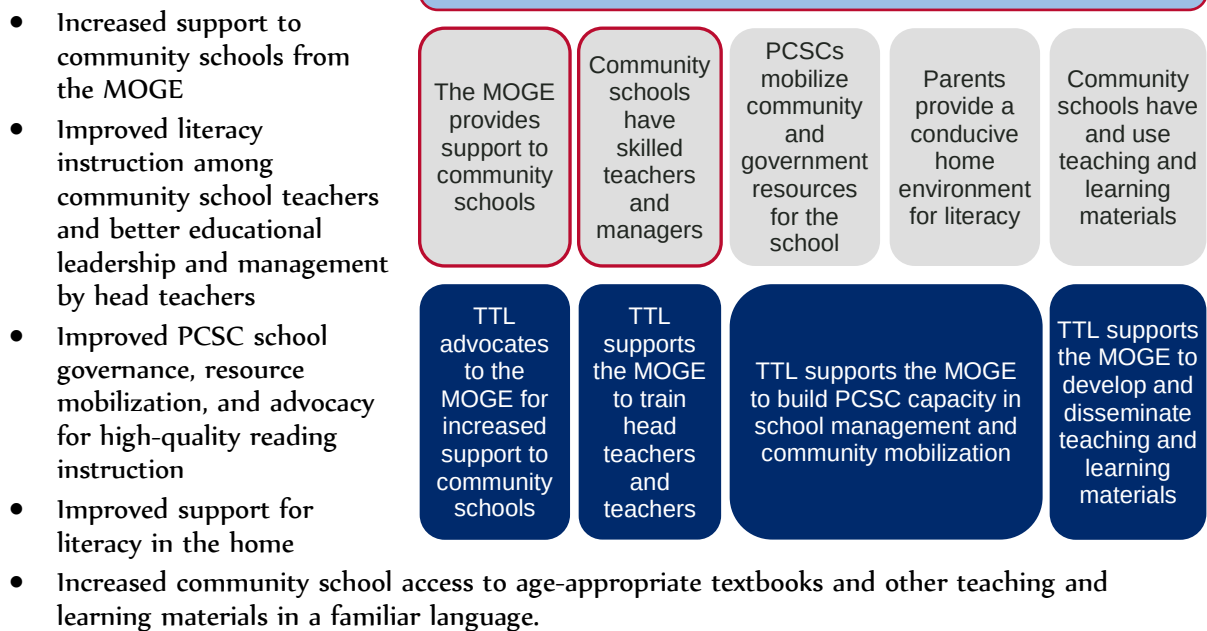
In recent years, addressing low reading performance has become a national priority. In 2013, the MOGE launched the new Primary Literacy Program, which incorporates a wide range of changes to literacy instruction for grades 1–4, including:

- Introducing a phonics-based approach to literacy instruction in the revised national curriculum (2013)
- Adopting a policy of familiar-language instruction in early grades for teaching reading in seven Zambian languages (replacing English as the medium of reading instruction) and beginning oral English lessons in grade 2 (2013)
- Developing and distributing new textbooks aligned to the Primary Literacy Program for grades 1–4 in the seven Zambian languages of instruction (2013–2017)
- Proposing benchmarks and targets for core early grade reading competencies (MOGE 2015a) that identify indicators for emergent readers (2015)
- Introducing the National Literacy Framework (MOGE 2013a), which specifies outcomes by each stage of literacy acquisition, provides a series of formative and summative assessments, and includes weekly grade 1 language schedules for the introduction and revision of letter sounds and blends (2013)
- Releasing a revised Zambian Languages Syllabus (MOGE 2013b) for grades 1–7 that specifies lesson topics and learning outcomes (2013).

2.2. TTL INTERVENTIONS

Exhibit 8: TTL Development Hypothesis

TTL's development hypothesis ([Exhibit 8](#)) was that community school learners would have improved reading as a result of the following changes (represented by the gray boxes):



TTL endeavored to achieve these results through a series of interventions (represented by the four dark blue boxes):

- MOGE capacity building
- Teacher training
- PCSC capacity building
- Teaching and learning material development and dissemination.

MOGE CAPACITY BUILDING

TTL worked with the MOGE through its existing structures and systems to reinforce its capacity to plan, support, manage, monitor, and evaluate community school progress toward improved education outcomes and to diffuse literacy and community school policy updates. TTL supported the MOGE to:

- **Incorporate reading assessment** into routine monitoring of community schools via [eEGRA Instruct](#), a proprietary software program designed by Education Development Center, Inc., for formative monitoring and feedback of reading performance at the school level (2014). As part of this activity, TTL provided a 4-day training course for provincial and district MOGE officials on administering eEGRA Instruct ([Exhibit 9](#)) and distributed 133 eEGRA Instruct-equipped netbooks to MOGE administrators, who were expected to train zonal in-service coordinators (ZICs) in their districts. TTL provided additional eEGRA Instruct and netbook technical support to MOGE officials as needed.
- **Form a community school steering committee**, which contributed to coordinating MOGE support to community schools across departments and with donors. The steering committee conducted two national community school symposia (2014 and 2015), revised the 2007 *Operational Guidelines for Community Schools* (OGCS) to improve the policy framework

governing them (2014), drafted a community school section for the National Policy on Education (in progress), and advocated for the inclusion of community schools in the Zambia National Budget, which occurred in 2016

- **Improve early grade reading strategy and policy** to provide a conducive framework for the teaching and learning of literacy in early grades. Through routine interaction with MOGE officials at all levels of the central ministry, TTL technical specialists advised on multiple strategies, processes, systems, curricula, and teaching and learning materials that support early grade reading, including the numerous changes released as part of the Primary Literacy Program and a Teacher Competency Framework. The framework, the first of its kind in Zambia, outlines the key skills, knowledge, and attributes necessary for community school teachers. This engagement further ensured that these updates considered the perspective of community schools and reached them during dissemination. In addition, MOGE officials were actively involved in all TTL research and evaluation activities to build interest and capacity to conduct high-quality data collection and reinforce best practices for education evaluation.

TEACHER TRAINING

TTL conducted training, designed in collaboration with the central MOGE, that cascades down through the MOGE teacher education structures from provincial, district, and zonal levels, ultimately reaching community school teachers. The TTL training listed below included variations on this cascade model, based on content and resource constraints. [Exhibit 9](#) summarizes the number of individuals who received TTL training at all levels of the cascade.

Exhibit 9: Individuals Who Received TTL Training

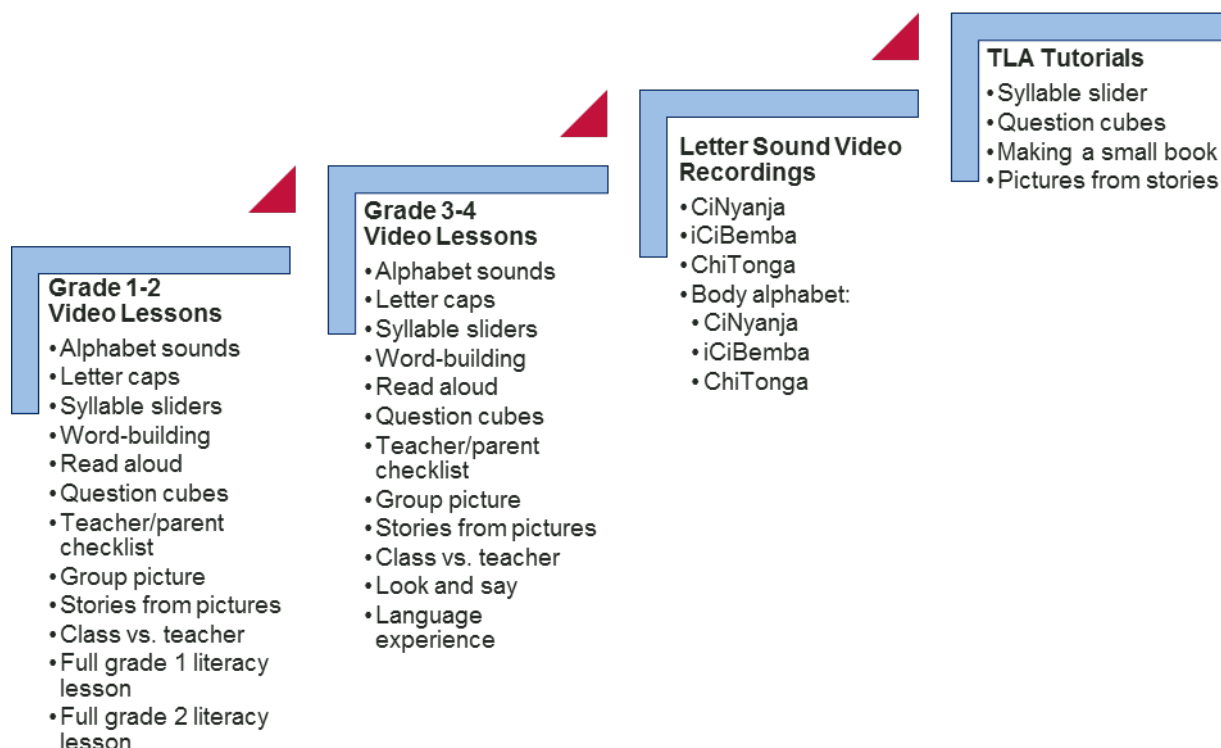
Training	2013	2014	2015	2016	Total
eEGRA Instruct	0	0	215	0	215
Quick Start Literacy Program	152	83	0	0	235
Educational Leadership and Management	1,558	631	0	0	2,189
Zonal Training: Writing	930	630	0	0	1,560
Zonal Training: Reading	0	67	0	0	67
Zonal Training: Alphabet Sounds	2,240	358	0	0	2,598
CPD on School-Based Assessment	504	4,481	144	0	4,985
Early Grade Reading Stepping Stone Program	0	0	2,752	2,617	5,369
Guiding Learners to Read and Classroom Coaching	0	0	0	5,157	5,157
Operational Guidelines for Community Schools	1,195	1,020	0	0	2,215
Community Mobilization to Support Literacy	1,167	1,279	11	0	2,446
Total	7,746	8,549	3,122	7,774	54,072

Source: TTL Monitoring Data

TTL and the MOGE conducted the following training events, from project inception:

- **The Quick Start Literacy Program** trained head teachers in literacy instruction basics through five modules: Comprehension, Fluency, Phonemic Awareness and Phonics (word decoding), Read Aloud (oral passage reading), and Vocabulary and Spelling (2013; 2 days).
- **Education Leadership and Management** training provided head teachers and PCSC chairs with skills in managing resources, information, and records; conducting and supervising school-based assessments; assessing effective teaching; developing school improvement plans; monitoring and evaluating school performance; and providing psychosocial support to vulnerable learners through counseling, environment, health, and hygiene education (2013–2014; 2 days).
- **Zonal Training and Teacher Learning Circles** included three 1-day modules as part of a CPD program: Reading (read aloud) and Writing in 2013 and Alphabet Sounds (letter sounds) in 2014. Zonal training participants were head teachers who were expected to train their teachers in the same material through “learning circles” at their schools.
- **CPD on School-Based Assessment**, for head teachers and teachers, focused on how to monitor learner progress and intervene accordingly. The workshop also reviewed teaching methods for literacy instruction and provided an orientation to the MOGE’s Primary Literacy Program (2014; 2-day training).
- **Early Grade Reading Stepping Stone Program** was loaded on Nokia 111 mobile phones, with one phone provided to each school. The [Stepping Stone](#) platform is proprietary software from Education Development Center, Inc., that provides video demonstrations of lessons, short quizzes, and other resources for teachers to use in preparing literacy lessons ([Exhibit 10](#)). In 2015, TTL conducted a first round of 2-day training workshops with district resource center coordinators, ZICs, and head teachers on how to use the program to prepare literacy lessons. Based on feedback, TTL held a second training round with additional content, including examples of full literacy lessons and videos of letter sounds, in 2016. The head teachers who participated were expected to train teachers at their schools, and the full Stepping Stone content was shared with all district resource center coordinators. At the request of the Curriculum Development Center, in 2016 TTL conducted a third training round with staff from the Center to continue building the MOGE’s base of teaching and learning materials for early grade reading.

Exhibit 10: Early Grade Reading Stepping Stone Program Training Content



- **Vernacular** included a series of literacy activities in the languages of instruction, delivered on touchscreen tablets running the Stepping Stone program, to build key reading skills. TTL conducted a pilot study of Vernacular in 10 community schools in 2015, comparing results to 10 other community schools that were given a traditional, paper-based version of the content and another 10 that received no additional tools. As part of the pilot, TTL trained 10 teachers (1 per school) to guide learners in using the tablets; the 10 teachers at comparison schools using the paper-based version received a parallel form of training.
- **Guiding Learners to Read and Classroom Coaching** targeted MOGE officials charged with supervising community school teachers: district resource center coordinators, ZICs, and zonal heads. TTL training provided these officials with coaching skills to help them mentor teachers in early grade reading instruction, including conducting basic classroom observation covering the core literacy instructional domains and providing post-observation feedback using appropriate language (2016; 4 days). Teams of district resource center coordinators were then tasked to develop training responses to the needs they found during observations, including generating their own training manuals and using the manuals to train two primary grade teachers from each community school in their district (2016; 2 days).

PCSC CAPACITY BUILDING

The PCSCs that manage community schools generally comprise parents, teachers, and prominent community members. TTL has built capacity of PCSCs through four primary activities, with the number of individuals trained summarized in [Exhibit 9](#):

- **Orientation on the OGCS** was provided through a 2-day training course for two members of each PCSC after the revised guidelines were complete (see “MOGE Capacity Building,” above).

The orientation gave an overview of the policy framework that governs community schools to better position the PCSCs to advocate with the MOGE on behalf of their schools (2013–2014).

- **Community Mobilization to Support Literacy**, a workshop to help PCSC members promote more supportive home literacy environments in their communities. Two members of each PCSC participated, after which they were expected to conduct community-level training to increase parents' awareness of the importance of education and how they could help their children learn to read (2013–2014, 2 days).
- **Design of school improvement plans by PCSCs** took the form of support from ZICs at planning workshops, as a follow-on to the 2013 and 2014 Education Leadership and Management training (see "Teacher Training," above). ZICs guided the PCSCs through a self-assessment process, including needs identification and planning for future improvement, and TTL supported the ZICs to help PCSCs create, finalize, or update their schools' implementation of improvement plans (2015–2016).
- **A literacy radio campaign** aired educational messages targeting parents of young learners on local radio stations. The campaign included eight programs that used short dramas, lectures, and dialogues in local languages to encourage and instruct parents on how to be more involved in their children's education and raise awareness about the importance of literacy (2015–2016).

TEACHING AND LEARNING MATERIAL DEVELOPMENT AND DISSEMINATION

TTL developed and disseminated more than 1 million low-cost, easily replicable supplementary reading materials and instructional resources to improve reading instruction in community schools in the six-province intervention area, as summarized in [Exhibit II](#). These materials have included:

- **A reading/learning kit for learners** with story cards, flashcards, short-story books (Maiden readers), and leveled CASAS readers¹ in the languages of instruction. An initial distribution took place in 2013–2014, with a final distribution in 2016.
- **Community library book boxes** were distributed to 150 community schools through a public-private partnership with TOTAL Zambia. The boxes included an array of books in English and the languages of instruction (2013–2014).
- **Instructional and management materials** included the 2012 *Zambian Basic Education Syllabi* and teachers' guides, such as the language schedule, attendance logs, enrollment forms, and daily and monthly continuous assessment booklets. Many of the materials disseminated as part of community school teacher training provided additional instructional resources, such as eEGRA Instruct: Literacy Activity Handbook for Teachers (2013–2016).
- **Technology-based instructional resources for teachers** included the video and audio content in the early grade reading Stepping Stone phone and Vernacular tablets.

Exhibit II: Teaching and Learning Materials Distributed

Type of Material	Total
Reading/learning kit for learners (CASAS and Maiden readers, story cards, flashcards)	420,743

¹ TTL readers are referred to by their publishing company: Maiden Publishing House and Center for Advanced Studies of African Society (CASAS).

Type of Material	Total
Instructional and management materials (attendance registers, continuous assessments, enrollment forms, training materials)	776,310
Technology-based instructional resources for teachers (phones, tablets)	2,809

Source: TTL Monitoring Data

3. EVALUATION DESIGN, METHODS, AND LIMITATIONS

This endline evaluation, designed as a learning and accountability tool for TTL and its stakeholders, integrated participatory and utilization-focused approaches with quantitative methodologies to engage key stakeholders in evaluation planning and data collection.

3.1. EVALUATION DESIGN

The evaluation team developed the initial evaluation design (including baseline, midline, and endline) in July 2012 through a consultative meeting with TTL and the MOGE's Directorate of Standards and Curriculum (including the Curriculum Development Center), Directorate for Planning and Information, district- and provincial-level officials, and the ECZ. At midline, the evaluation team developed an implementation plan with the TTL project team and the ECZ, detailing adaptations to the July 2012 design to respond to changes in project activities, challenges and successes encountered during the baseline, and updated USAID guidance for measurement of [USAID Education Strategy](#) Goal 1. An endline implementation plan specified the design for the endline evaluation phase and ensured compliance with the second edition of USAID's EGRA Toolkit (RTI International 2015), but design changes were not required.

In keeping with USAID-recommended approaches for measuring reading gains, the evaluation uses an independently pooled, repeated cross-sectional design. The evaluation does not include a counterfactual, given the broad coverage of TTL's interventions and national rollout of the revised literacy curriculum. Key evaluation questions 1 and 2, listed in [Exhibit 12](#), are answered through comparison with (1) baseline data collected before project implementation to measure the total amount of change over the 5-year project period, and (2) midline data to facilitate understanding of the rate of that change. The answer to evaluation question 3 is based on a comparison with midline data only, as baseline data were collected using a different instrument (see ["Classroom observation protocol" in Section 3.2](#)). Data collection for all three phases occurred at the same point in the school year, as required by the cross-sectional design. [Annex 5](#) provides a list of evaluation questions by data source and indicator.

SAMPLE DESIGN

At each evaluation phase, sampling used a two-stage design, with an additional intermediary stage applied in schools with more than a single grade 2 class. To provide provincially disaggregated estimates of EGRA results, target sample sizes were calculated for the provincial level and separate samples drawn for each province.

- **Stage 1:** A random sample of schools was drawn for each province using probability proportional to size with schools serving as clusters. In Lusaka Province, the sample was additionally stratified by rural and urban areas.²
- **Intermediary stage:** The vast majority of schools sampled had only one grade 2 section (“stream”). For schools with multiple grade 2 streams, one stream was selected through simple random sampling.³
- **Stage 2:** A sex-stratified simple random sample of grade 2 learners was drawn at each school (or from the selected class, for schools with multiple grade 2 streams).

Consistency in the sampling methodology across phases reflects the cross-sectional design. See [Annex 3](#) for more details on the sample, including target sample size and power calculations. [Exhibit 15](#) in Section 4 presents the achieved endline sample; baseline and midline samples are in [Annex 3](#).

The evaluation was designed to respond to the TTL performance monitoring plan, which specifies EGRA measurement by province. In Zambia, each province has an official language of instruction:

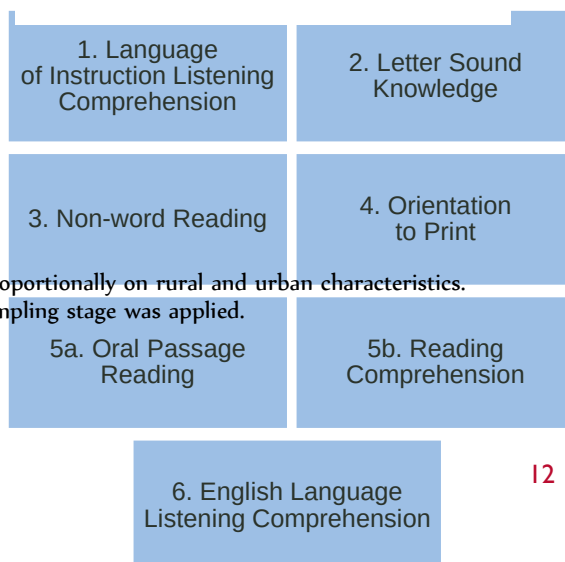
- **ChiTonga:** Southern
- **CiNyanja:** Eastern and Lusaka
- **iCiBemba:** Central, Copperbelt, and Muchinga

Three provinces in the TTL intervention area—Central, Lusaka, and Muchinga—have more than one language of instruction. In Central Province, the primary language of instruction is iCiBemba, with four districts where ChiTonga is used. The primary language of instruction in Muchinga Province is iCiBemba, with one district where CiNyanja is used. The evaluation samples drew only from iCiBemba-speaking districts in Central and Muchinga provinces. The primary language of instruction in Lusaka Province is CiNyanja, but several schools use ChiTonga; only CiNyanja-speaking schools were sampled. *Consequently, EGRA results from these provinces reflect only the population of primary language of instruction community schools.*

3.2. DATA COLLECTION METHODS AND DATA COLLECTORS

In October 2016, the evaluation team conducted a school-based survey using the following tools and methods: EGRA, classroom observation, head teacher questionnaire, and a self-administered questionnaire for MOGE District Education Boards. Brief descriptions of these tools and methods are below, with additional detail EGRA, the classroom observation protocol, and the community school head teacher questionnaire in [Annex 3](#), including a description of differences in tools across the three phases. The complete set of tools is provided in [Annex 7](#).

EGRA: The TTL EGRA includes a learner interview to capture key demographic information about respondents and an assessment component that measures four core literacy competencies—alphabetic principle awareness and decoding (phonics), oral language, reading fluency, and reading comprehension—through seven tasks ([Exhibit 13](#)).



² Lusaka is the only province with enough urban schools to sample proportionally on rural and urban characteristics.

³ See [Exhibit 15](#) for the number of schools to which this additional sampling stage was applied.

EGRA was administered in three languages—ChiTonga, CiNyanja, and iCiBemba—corresponding to the official language of instruction of each province. The assessment portion of the TTL EGRA aligns with the Zambia EGRA used by ECZ and other USAID literacy projects; the instrument has been equated across evaluation phases to enable comparison of results (see [Annex 4](#)). EGRA tasks 2, 3, 5a, and 5b are comparable across all three phases; tasks 1, 4, and 6 are comparable only between midline and endline.

Classroom observation protocol: TTL’s classroom observation protocol contained a structured classroom observation followed by a closed-ended interview questionnaire on demographics and teacher practice, administered face-to-face with the observed teacher. Classroom observation examined 26 distinct indicators of early grade literacy lesson quality and instructional delivery techniques. Data collectors observed lessons and recorded fulfillment of the 26 criteria in 3-minute intervals over a period of 60 minutes (based on the MOGE standard that the primary grade literacy lesson should last 1 hour per day), creating a maximum of 20 intervals (not all lessons lasted the full hour). These criteria were grouped into six domains, with an additional section covering lesson delivery methods (see [Exhibit 14](#)) that combine criteria into broader categories that reflect the core pillars of early grade literacy lessons. These domains correspond to the Zambia EGRA and the TTL intervention model.

The classroom observation tool used at baseline was poorly suited to TTL’s intervention model, requiring adaptation of the tool; consequently, comparisons are available only between midline and endline. Small improvements to the tool at endline mean there are a few criteria between midline and endline that are not comparable.

Exhibit I4: Classroom Observation Domains

Part 2A: Literacy Content (20 criteria)	
Domain I: Orientation to Print	Domain IV: Reading Comprehension
Domain II: Phonemic Awareness and Phonics	Domain V: Oral Language
Domain III: Passage/Story Reading (fluency)	Domain VI: Writing
Part 2B: Lesson Delivery (6 criteria)	

Community school head teacher questionnaire: A face-to-face interview questionnaire was administered to a representative of every school in the sample. The questionnaire captured data related to the school profile, the head teacher's background, school funding and resources, MOGE support, instructional practices, educational leadership and management, the PCSC, and learner absenteeism and attrition. To improve validity of respondents' answers pertaining to TTL teaching and learning material and training support received, each data collection team had a complete set of TTL teaching and learning materials, which they showed to respondents during question probing to confirm that the school had actually received the materials. Where possible, presence of TTL-disseminated materials was physically verified. A core suite of questions related to MOGE support and the school and head teacher's profiles are comparable across all three evaluation phases; additional questions specific to the evaluation question were included where appropriate.

MOGE self-administered questionnaire: Provided to the MOGE District Education Board Secretary during the evaluation team's initial courtesy call in each district, the questionnaire asked specific questions about how the MOGE had supported the community schools during the current and previous academic years. These data were triangulated with community school head teachers' perceptions of support. Because this method was not used at baseline, at midline the District Education Board Secretaries were asked to report on support in 2012 as well. To provide for accuracy in triangulating midline and endline MOGE support between this source and the community school head teacher questionnaire, this tool targeted only schools in the sample for that particular evaluation phase. The District Education Board Secretary had responsibility for ensuring that the relevant district MOGE officers were consulted for each question. For example, the District Education Board Secretaries were to consult with finance officers for information pertaining to direct financial support provided to community schools in the sample, with human resources officers for the number of teachers deployed to the schools sampled, and so on. Offices had up to 1 month to collect the requested information. Of the 43 districts in which data collection took place, 42 returned the survey (98 percent), providing information on 98 of the 105 schools surveyed at endline (93 percent).

DATA COLLECTORS

All data collectors were MOGE officials, University of Zambia School of Education students, or TTL staff. Each data collector received 1 week of training for the evaluation tool they would administer (EGRA, classroom observation protocol, or head teacher questionnaire) and participated in sessions on data quality and evaluation ethics. Training sessions were held in September 2016, immediately prior to data collection. Training on assessor bias mitigated against risk of TTL staff bias. The evaluation team conducted inter-rater reliability scoring for all potential EGRA assessors and classroom observation protocol administrators during the sessions to ensure that all data collectors were administering the tools consistently. Descriptions of the inter-rater reliability testing and results are provided in [Annex 3](#).

3.3. ANALYSIS

The evaluation team analyzed the data in Stata version 14 and systematically archived all output files to facilitate replicability. The team first analyzed all data using descriptive statistics, with particular attention to group distributions, before proceeding to bivariate and inferential techniques. All data were analyzed for similarities and differences between males and females. Head teacher questionnaire and classroom observation data were compared at the provincial level, but the sampling was not designed to detect significant differences at that level. Similarly, provincial-level EGRA results were disaggregated by sex, but the sample was designed to provide accurate estimates only at the sex-aggregated provincial level. ***All differences that are statistically significant at the 5 percent level are noted as such in the text. Unless specifically noted, differences should not be construed as statistically significant.*** Regression analysis using ordinary least squares and controlling for the sample design characteristics was conducted following examination of correlation and bivariate analysis.

Analysis of EGRA data included zero score analysis because, based on the time frame and reading levels observed in 2012, it was expected that a notable proportion of learners would struggle to complete most EGRA tasks.

3.4. LIMITATIONS

The evaluation design was not able to statistically attribute changes to the interventions; TTL's universal coverage of all community schools in six provinces renders the creation of a comparison or control group impossible. Limitations in the EGRA instrument mean that results cannot be compared across languages. As with all survey questionnaires, recall issues among head teachers and teachers may limit data accuracy. Additionally, the evaluation encountered the following challenges in instruments and data across the three phases: the newness of the EGRA and classroom observation instruments in Zambia in 2012 and the use of MOGE personnel who were not experienced data collectors with these instruments (baseline); premature school closings in Muchinga Province (midline); and no 2012 data for EGRA task 4, orientation to print and task 6, English language listening comprehension, as well as a lack of comparable data for task 1, language of instruction listening comprehension (baseline). Because the baseline classroom observation protocol did not include the specific classroom practices promoted by TTL, a new tool was introduced at midline and used again at endline; therefore, midline and endline classroom observation data are not comparable to baseline. [Annex 3](#) presents a full description of the causes and implication of these limitations.

4. DESCRIPTION OF THE ENDLINE SAMPLE

This section outlines key characteristics of the endline sample to (1) gauge sample comparability across the three evaluation phases and (2) provide context for the findings. [Exhibit 15](#) presents the overall sample for the endline phase of this evaluation. Sample targets were met or surpassed in all provinces at endline. Sample summary tables for baseline and midline are provided in [Annex 3](#).

The remainder of this section presents descriptive statistics for the endline data, noting similarities and differences with baseline and midline data when relevant. Overall, the samples are comparable across the three phases.

Exhibit 15: Endline Evaluation Sample Overview

Province	Schools: Sample Stage 1*	Learners Participating in EGRA: Sample Stage 2			Teachers Participating in Classroom Observation			Head Teacher Questionnaire Respondents**		
		Male	Female	Total	Male	Female	Total	Male	Female	Total
Central	17	154	160	314	8	9	17	14 (2)	3	17 (2)
Copperbelt	17 (2)	133	150	283	5	12	17	9 (3)	8 (1)	17 (4)
Eastern	17	156	151	307	8	9	17	12 (2)	5	17 (2)
Lusaka	19 (2)	150	146	296	6	13	19	12 (5)	7	19 (5)
Muchinga	17	148	148	296	11	6	17	12 (1)	5	17 (1)
Southern	18	140	162	302	12	6	18	15 (3)	3	18 (3)
Total	105 (4)	881	917	1,798	50	55	105	74 (16)	31 (1)	105 (17)

* Parentheticals denote schools with more than one grade 2 class.

** Parentheticals denote participants with other titles.

4.1. CHARACTERISTICS OF THE SCHOOLS IN THE SAMPLE

LEARNER ENROLLMENT

Across the 105 community schools sampled at endline, average total school enrollment was 283, with wide variation (49–1,311), as [Exhibit 16](#) illustrates. These enrollment numbers are consistent with midline and baseline sample schools and with TTL monitoring data, which found an average total enrollment of 219 across 2,001 community schools in the project's catchment area ([Exhibit 7](#)).

[Exhibit 17](#) shows average school enrollment by province and by sex, with no significant differences by sex overall or across provinces. The visual differences by province were skewed by a few large schools in Copperbelt and Lusaka provinces. No differences in school enrollment existed between male and female learners, which aligns with MOGE data on enrollment by sex in primary grades in Zambia (MOGE 2016a).

Exhibit 16: Total School Enrollment

(n=105 community schools)

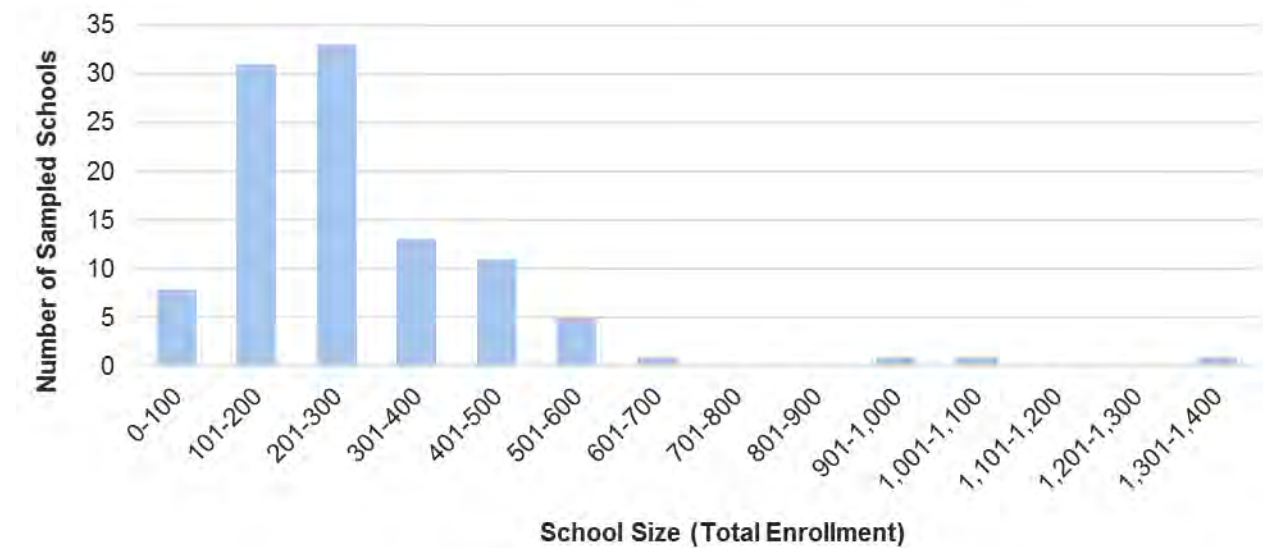
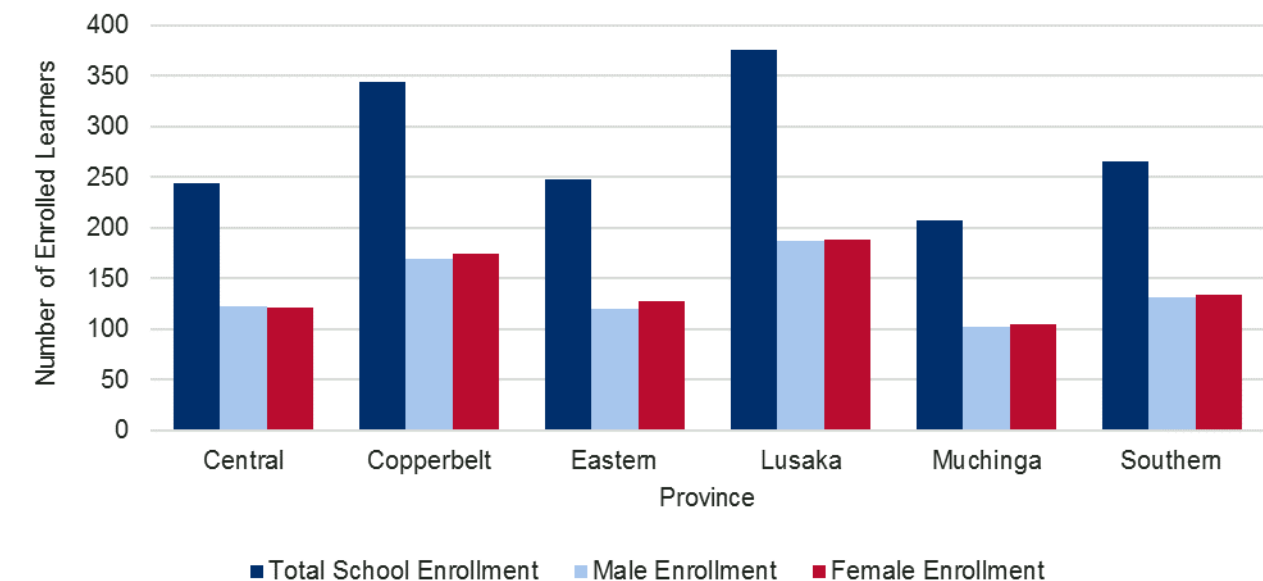


Exhibit 17: Average Total School Enrollment by Province and Sex

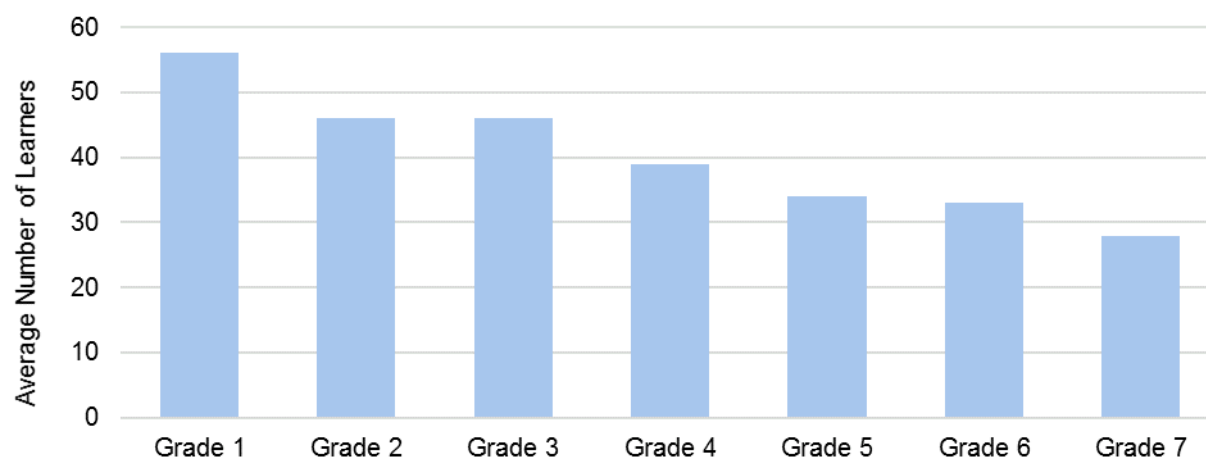
(n=105 community schools)



Enrollment decreased as grade level increased across the 105-school sample (see [Exhibit 18](#)), indicating that a substantial number of learners were either failing to progress or were exiting community schools as they progressed to higher grades. Overall, the sample indicated gender parity across grades.

Exhibit 18: Average Enrollment by Grade

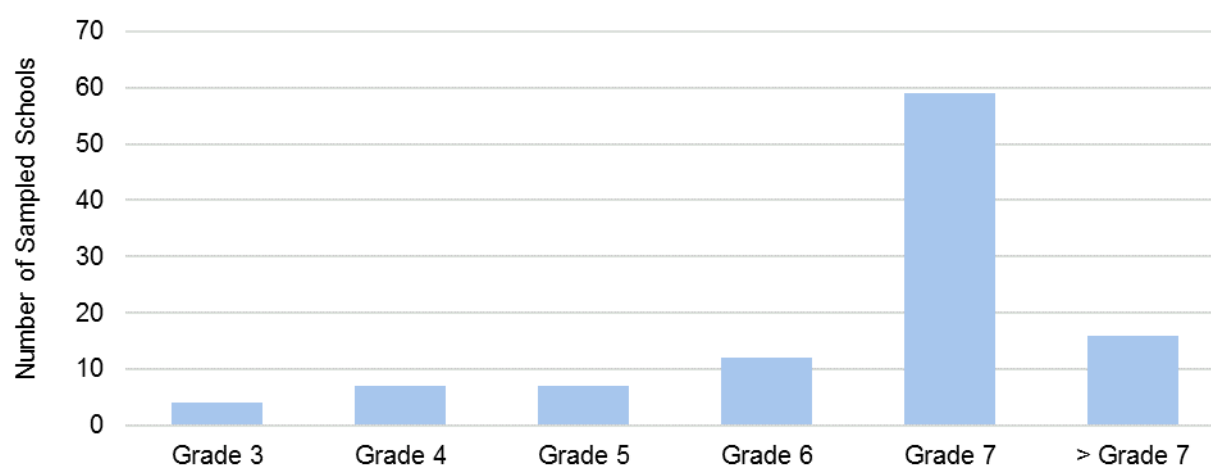
(*n=105 community schools*)



[Exhibit 17](#) and [Exhibit 18](#) present averages only among the schools actively teaching those grades. The majority of sampled schools taught grades 1–7, as shown in [Exhibit 19](#). All the schools teach at least up to grade 3, and some continue beyond grade 7. This is similar to the schools in the baseline and midline samples.

Exhibit 19: Highest Grade Taught in Sampled Schools

(*n=105 community schools*)



TEACHER AND SCHOOL PERSONNEL

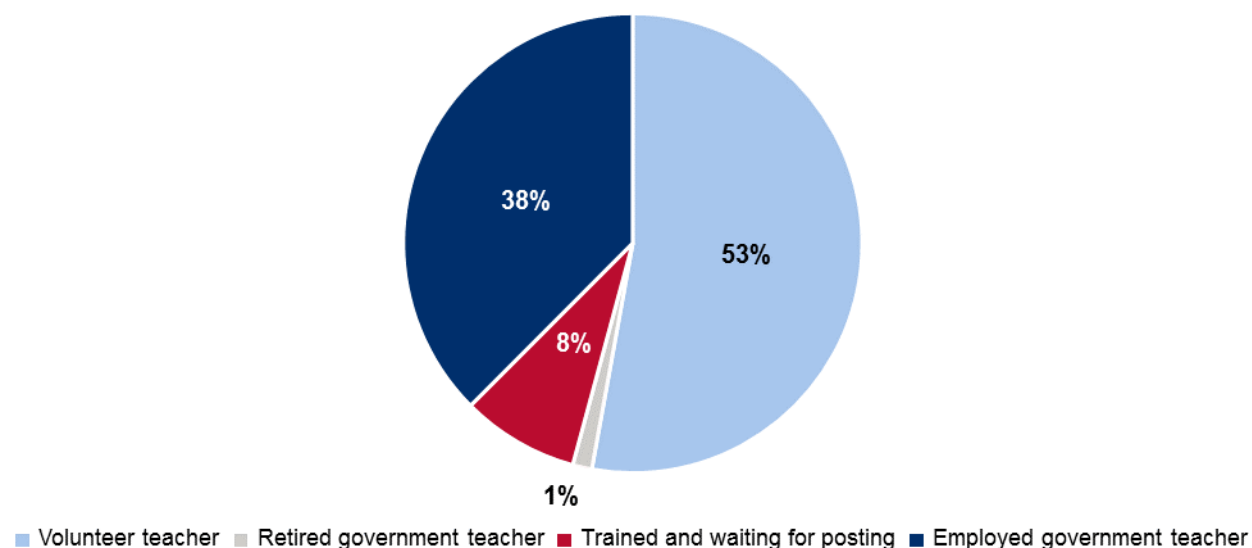
In the 105 community schools sampled at endline, the average number of teachers per school was 6.32, almost evenly divided between male and female teachers. However, the average number of male teachers across schools is 61 percent, indicating that smaller schools are more likely to have more male teachers. Copperbelt and Lusaka provinces have the highest mean and median numbers of teachers (19.53 and 9.68 as means, and 8 and 6 as medians, indicating that a few large schools are skewing the average upward); the other provinces had means ranging from 3.53–4.88.

[Exhibit 20](#) shows the employment status of community school teachers among the sampled schools, indicating that the majority (53 percent) are volunteer teachers. The percentage of government

employed teachers (38 percent) is an increase from the midline (26 percent). The average proportion of volunteer teachers across schools is 65 percent, again indicating that teachers in smaller schools are different than those in larger schools—in this case, more likely to be volunteer teachers.

Exhibit 20: Teacher Employment Status in Sampled Schools

(*n=105 community schools*)



The average learner-to-teacher ratio is 60:1 across all 105 schools, with a range of 18–315 learners per teacher, in line with the midline sample but higher than the baseline figure (47:1 average learner-to-teacher ratio). The endline school with 315 learners per teacher was in Muchinga Province, due to the school's recent loss of two teachers. Average learner-to-teacher ratios at endline vary by province, from 43:1 in Copperbelt to 80:1 in Muchinga.

AGE AND LOCATION OF SCHOOLS AND PRESENCE OF PCSC

The average length of time sampled schools have been in operation is 14.7 years (similar to the midline), with limited variation among provinces. Seventy-eight percent of sampled schools are in rural areas, also comparable to midline. Sampled schools varied widely within and across provinces in the distance to the nearest MOGE District Education Board Secretary office (see [Exhibit 21](#)), with schools in Copperbelt and Lusaka provinces having shorter distances, on average.

Exhibit 21: Distance to Nearest District Education Board Secretary Office

(*n=105 community schools*)

MOGE District Office	Lusaka	Copperbelt	Muchinga	Eastern	Southern	Central
Mean Distance (in kilometers)	24.6	29.2	58.2	64.9	68.6	72.1

All but two of the sampled schools (98 percent) have a PCSC, and of those that do, 55 percent of PCSCs report meeting monthly and 38 percent report meeting quarterly. Ninety-three percent of these committees receive academic reports from the school. Three-quarters of these PCSCs build and maintain school facilities, and about half monitor teacher and learner attendance and observe classes.

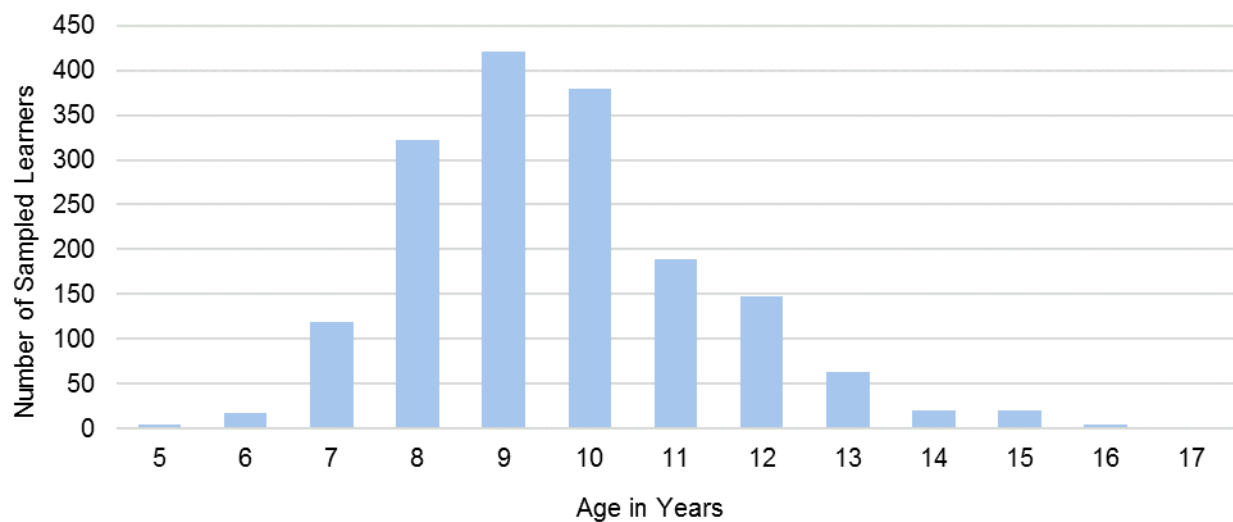
4.2. CHARACTERISTICS OF LEARNERS IN EGRA SAMPLE

SOCIODEMOGRAPHIC CHARACTERISTICS

The EGRA sample was designed to assess an equal proportion of male and female learners. A total of 1,798 learners were assessed at endline, almost equally divided by sex (51 percent female and 49 male). [Exhibit 22](#) shows the distribution of age among sampled learners in grade 2, with the mean age being 9.63 years. There is no substantial variation by province or sex, and age is comparable to the baseline and midline samples. However, 19 percent of the sample comprised learners 12 years of age or older, 4 years older than the official age (8 years) for grade 2 children in Zambia.

Exhibit 22: Age of Sampled Grade 2 Learners

(*n=1,798 learners*)



In an effort to gauge the socioeconomic status of participants, learners were asked to identify assets owned by their household from a basket of 10 items appropriate to the Zambian context. As shown in [Exhibit 23](#), learners in the EGRA sample generally came from households owning few of these items, indicating a generally vulnerable population of learners, with an overall sample median of three items per household. There was limited variation across provinces, with patterns similar to the midline sample.

Exhibit 23: Distribution of Sampled Learners by Number of Household Assets

(*n=1,798 learners*)



LANGUAGE CHARACTERISTICS

Learners were tested in the official MOGE language of instruction. In most cases, there is a single language of instruction throughout a province. However, Zambia has substantial linguistic diversity; not all sampled learners spoke the official language of instruction at home or in their schools (see [Exhibit 24](#)). Overall, 93 percent spoke the language of instruction at their school and 80 percent spoke the language of instruction at home; these numbers are consistent with the midline sample. As in the midline sample, there is variation across provinces, with Eastern Province having the lowest percentage of learners speaking the language of instruction at school and at home, and four provinces (Central, Copperbelt, Lusaka, and Muchinga) having differences of 15 percent or higher between those speaking the language of instruction at school and at home (indicating provinces where schools' catchment areas likely exhibit greater linguistic diversity).

Exhibit 24: Sampled Learners Who Speak the Language of Instruction and EGRA

(*n=1,798 learners*)

Province	At School	At Home
Central	97%	80%
Copperbelt	94%	78%
Eastern	74%	73%
Lusaka	98%	83%
Muchinga	92%	72%
Southern	100%	93%
TOTAL	92%	80%

ATTENDANCE PATTERNS OF LEARNERS

Attendance in the grade 2 class sampled, based on physical counts conducted by data collectors during school visits and compared to enrollment numbers reported by the head teacher, averaged 60 percent on the day of data collection (see [Exhibit 25](#)). Attendance varied by province, ranging from 52 percent of learners attending class on the day of data collection in Lusaka and Southern provinces to 66 percent in Central and Eastern provinces. There was little overall variation by sex (62 percent attendance for girls and 60 percent for boys), with a similar pattern across provinces, except for Muchinga. The overall rate of attendance on the day of data collection was 10 percentage points lower than at the midline, with the biggest differences in attendance rates between endline and midline in Copperbelt (21 percentage points lower at endline). Eastern, Lusaka, and Muchinga exhibited drops of 9 to 13 percentage points. Central and Southern provinces showed similar attendance rates. There is no data to explain these differences.

Exhibit 25: Grade 2 Attendance on Day of School Visit as Percentage of Total Enrollment
(*n=105 community schools*)

Province	Total	Male	Female
Central	66%	66%	69%
Copperbelt	54%	52%	56%
Eastern	66%	64%	68%
Lusaka	52%	51%	56%
Muchinga	61%	68%	62%
Southern	61%	56%	64%
TOTAL	60%	60%	62%

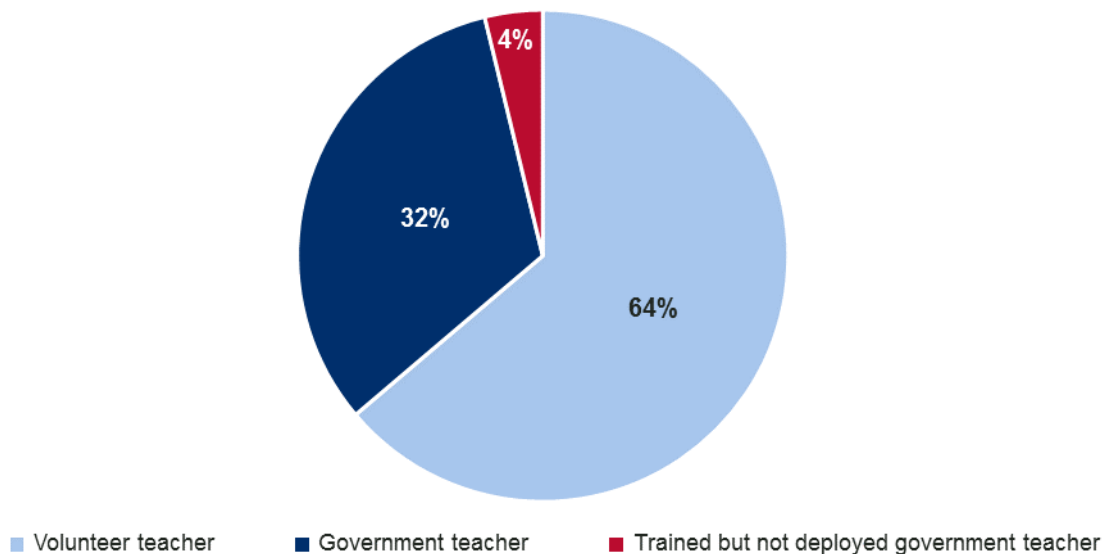
Among sampled learners (*n=1,798*), 83 percent reported attending school every day, a slight increase from the midline. Eleven percent of learners reported attending 4 out of 5 days per week, 4 percent reported attending 3 days, 0.8 percent reported attending 2 days, and 0.4 percent reported attending 1 day per week.

4.3. CHARACTERISTICS OF TEACHERS PARTICIPATING IN LITERACY LESSON CLASSROOM OBSERVATION

The evaluation team observed a total of 105 literacy lessons (one observation per school). Of the teachers observed, 48 percent were male and 52 percent were female, similar to the distribution at midline and with little variation among the provinces. The average age of teachers was 33.1 years (range 17–64).

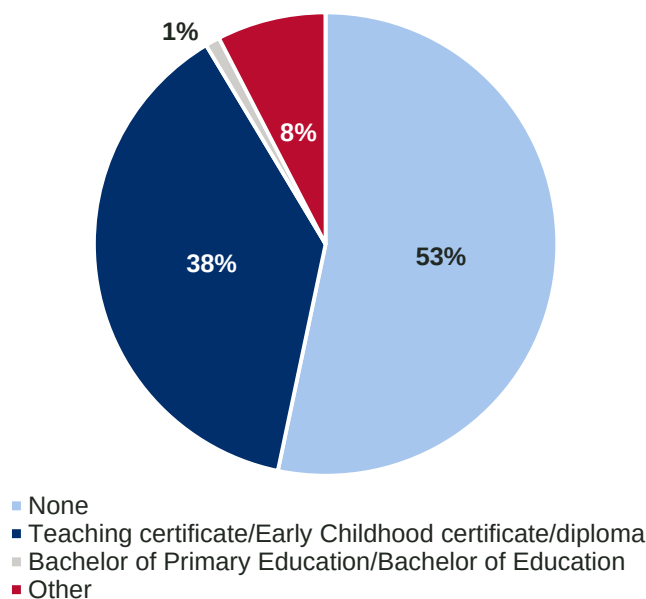
Sixty-four percent of observed teachers were volunteer teachers, as illustrated in [Exhibit 26](#). The percentage of government teachers among those observed was higher than at midline (32 percent, compared to 12 percent), and these government teachers were significantly more likely to be female than the volunteer teachers. Central and Lusaka provinces had the fewest government teachers among those observed. Of those observed, 20 percent were head teachers, 11 percent were deputy head teachers, and 9 percent were senior teachers.

Exhibit 26: Employment Status of Observed Teachers
(*n=105 teachers*)



The majority of teachers (71 percent) had passed the grade 12 exam, 9 percent had passed the grade 9 exam, and 9 percent had completed either grade 10 or grade 11. [Exhibit 27](#) illustrates the professional qualifications of the observed teachers, indicating that only 39 percent had any formal professional training. The average number of years teaching among those observed was 6.26 (median 5 years, range 0–18), with 4.53 (median 2 years, range 0–16) at the present school. These numbers were generally 2–3 years more than the averages at midline (which took place 2 years before).

Exhibit 27: Highest Professional Qualifications Achieved by Observed Teachers
(*n=105 teachers*)



4.4. CHARACTERISTICS OF COMMUNITY SCHOOL HEAD TEACHER RESPONDENTS

The community school head teacher questionnaire was designed to be administered to the head teacher, but in cases where a head teacher was unavailable, a deputy was allowed to answer on the head

teacher's behalf. In 88 of the 105 schools sampled, the head teacher served as the respondent. This section reports the qualifications of the head teacher of the school, not of the respondent.

Overall, head teachers at community schools have higher academic qualifications than classroom teachers observed, with more having passed the grade 12 exam (78 percent of respondents, compared to 70 percent of the classroom teachers observed). Fifty-seven percent had some kind of teaching certificate, degree, or diploma, and 32 percent had no pre-service training, an improvement from the midline. Head teachers were generally male (71 percent), similar to the baseline and midline findings. Although the pattern of more male than female head teachers is consistent across all regions, the percentages are much higher in Central and Southern provinces (82 and 83 percent), and much lower in Copperbelt (53 percent).

Community school head teachers had worked an average of 13.7 years as a teacher and 5.6 years as a head teacher at the sampled community schools. Forty-six percent of head teachers were supported by the government. Of the head teachers who were not government teachers, the majority were hired and paid by the PCSC.

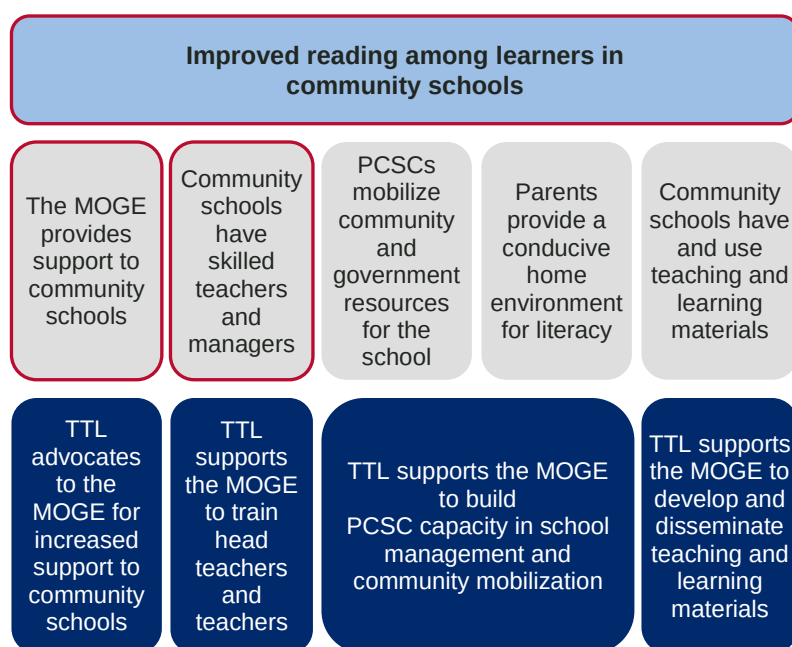
5. FINDINGS

The endline evaluation findings provide an understanding of changes since baseline in three areas of the development hypothesis, outlined in red in [Exhibit 28](#):

1. Reading among community school learners, represented by the top box (evaluation question 1)
2. MOGE support to community schools, represented (evaluation question 2)
3. Literacy instruction among community schools represented by (evaluation question 3).

The findings are presented below and organized by the three evaluation questions listed in [Exhibit 12](#).

Exhibit 28: TTL Development Hypothesis



GLOSSARY OF STATISTICAL TERMS

Effect size	A measure of whether gains are practically meaningful by indicating the size of the change on a standard scale. Effect size is a measure of <i>practical</i> significance that complements measures of statistical significance.
Mean	The average of the numbers. The mean is a calculated "central" value of a set of numbers.
Median	The point at which 50 percent of cases are above and 50 percent are below.
p-value	A measure of statistical significance ranging between 0 and 1. Lower values are more significant, meaning the observation is less likely to have occurred by chance. This evaluation reports findings as significant at the 5 percent threshold, indicated by a p-value equal to or less than 0.05.
Statistical significance	A high degree of confidence that differences are real and not simply a result of random chance due to sampling.
Zero score	The percentage of learners who did not complete any items correctly on an EGRA task; previous EGRAs in Zambia and across Sub-Saharan Africa have shown a high proportion of learners scoring zero. Consequently, reductions in these scores are highly positive, indicating that learners are making the critical first step in developing the skill being measured.

5.1. READING ACHIEVEMENT AMONG COMMUNITY SCHOOL LEARNERS

Findings in this section are based on the endline EGRA data and, where applicable, comparison to midline and baseline data. This section presents results graphically. [Annex 1](#) provides comprehensive summary tables for all EGRA tasks by province and language, including confidence intervals and disaggregation by sex.

[Exhibit 29](#) outlines the four Zambian reading performance benchmarks (non-reader, pre-emergent reader, emergent reader, and reader) for each EGRA task. These benchmarks are used in this evaluation to track the proportion of learners performing at each of the four levels over time. The MOGE stipulates benchmarks for three tasks: non-word reading (task 3), oral passage reading (task 5b), and reading comprehension (task 5b) in the draft document, *Proposing Benchmarks and Targets for Early Grade Reading and Mathematics in Zambia*, which the MOGE and ECZ drafted with support from USAID, through RTI International, following the 2014 Grade 2 National Assessment (MOGE 2015a). The evaluation team provided parallel classifications for the other tasks in the Zambia EGRA for consistency. Additionally, the evaluation team has supplemented the benchmarks with two additional categories: “non-readers” are defined as those learners who are unable to perform any items correctly on a task (those scoring zero); “pre-emergent” are defined as those learners scoring above zero but below the “emergent” benchmark (the lowest defined MOGE benchmark). The distinct subset of “non-readers” is important, because their performance implies different instructional needs.

Exhibit 29: Performance Categories for EGRA Tasks

Task	Non-reader	Pre-emergent Reader	Emergent Reader	Reader
Task 1: Zambian language of instruction listening comprehension	0 questions correct	1 question correct	2–3 questions correct	4–5 questions correct
Task 2: Letter sound knowledge	0 letters per minute	1–15 letters per minute	16–40 letters per minute	41+ letters per minute
Task 3: Non-word reading	0 words per minute	1–14 words per minute	15–29 words per minute	30+ words per minute
Task 4: Orientation to print	0 questions correct	1 question correct	2 questions correct	3 questions correct
Task 5a: Oral passage reading	0 words per minute	1–19 words per minute	20–44 words per minute	45+ words per minute
Task 5b: Reading comprehension	0 questions correct	1 question correct	2–3 questions correct	4–5 questions correct
Task 6: English language listening comprehension	0 questions correct	1 question correct	2–3 questions correct	4–5 questions correct

Note: The reading categories (“bins”) for tasks 3, 5a, and 5b are taken from the MOGE benchmarks for early grade reading (MOGE 2015a). The categories for tasks 1 and 6 are aligned to task 5b because the tasks are scored in a similar manner and on the same scale. Task 4 has only three questions; a separate bin is presented for each possible total score

As a whole, community school learners have shown significant improvements in reading skills since baseline, with a greater proportion achieving MOGE-defined benchmarks.

Exhibit 30 presents the proportion of learners estimated to be performing at each benchmark for four EGRA tasks (2, 3, 5a, and 5b) at baseline, midline, and endline across the six TTL intervention provinces and three languages of instruction. Examining these results reveals a substantial increase from baseline to endline in the proportion of learners at the emergent reader level:

- Letter sounds (task 2): 4 percent to 19 percent
- Non-word reading (task 3): 2 percent to 10 percent
- Oral passage reading (task 5a): 2 percent to 9 percent.

For reading comprehension (task 5b), there was a substantial upward trend across all three phases of the evaluation at the pre-emergent reader level.

Overall, the results indicate a large shift toward higher performance categories and a commensurate strong decrease in the proportion of non-readers across all four tasks. This decrease in non-readers is a highly positive development that is particularly evident when compared to the baseline, when 9 out of 10 learners could not read a single word. Although the proportion of learners performing at the emergent and reader benchmarks remained low at endline against the MOGE's targets for 2019,⁴ this progress is a necessary precursor. The strong decline in non-readers represents a qualitative shift: the majority of learners are now beginning to grasp basic literacy concepts and are taking the first steps in learning to read. In addition, ***TTL exceeded its own 2016 target, with 5.7 percent of community school learners reading 25 or more words correctly per minute.***⁵

The improvement in phonics, decoding, and reading fluency is due not only to a decrease in non-readers, but also to an increase in the proportion of learners achieving at the higher level of emergent reader, indicating that the improvement is happening along the full distribution. In other words, performance in these skills has improved across the spectrum, not just at the lower end. The improvements from baseline to midline reveal that gains in TTL's first 2 years were in the pre-emergent and emergent categories. Increases between midline and endline are seen mainly in the higher, emergent category for decoding and reading fluency, while for reading comprehension growth is still primarily at the pre-emergent level.

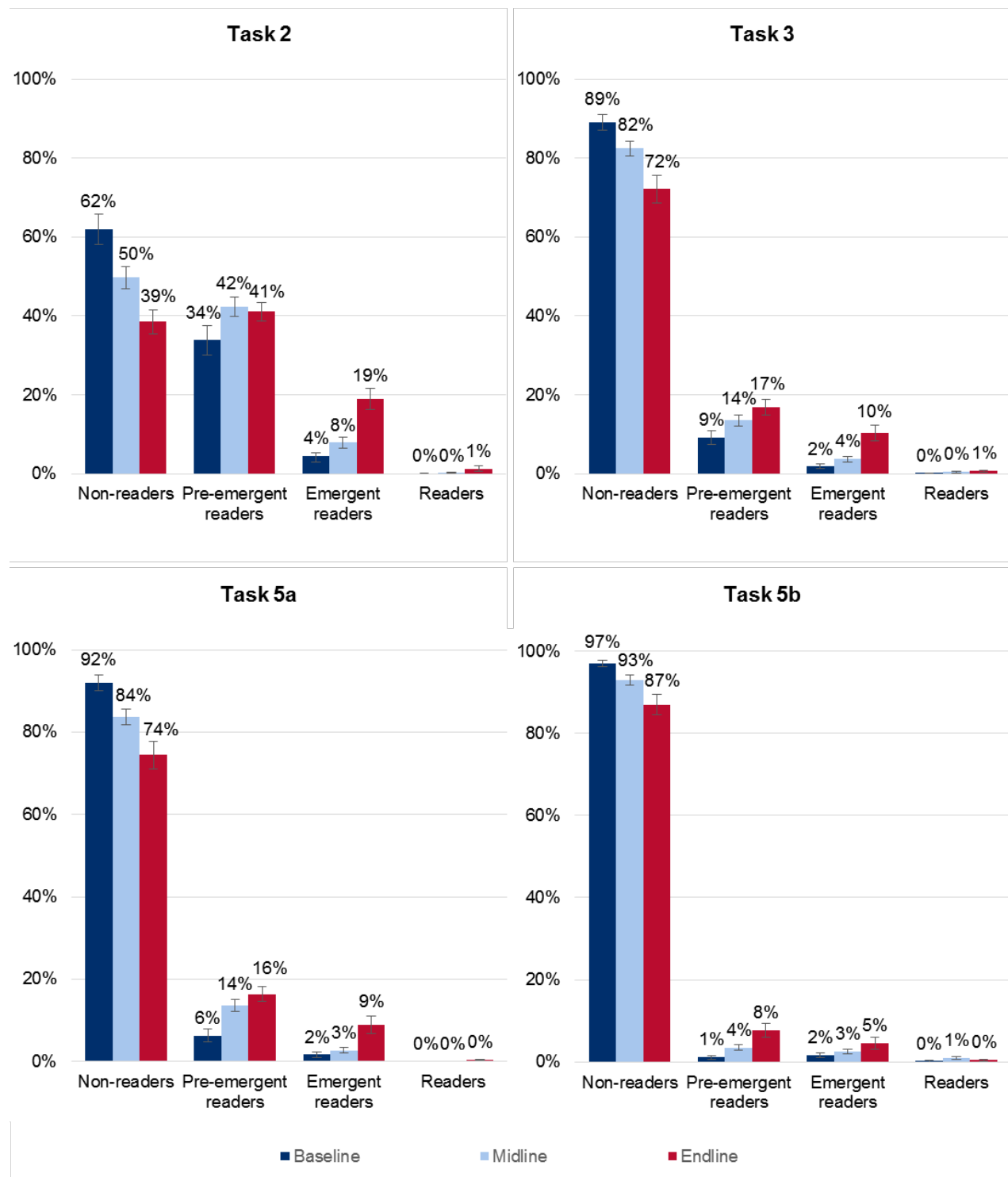


At baseline, 9 out of 10 learners could not read a single word; there have been substantial increases in the percentage of learners performing at the pre-emergent and emergent benchmarks.

⁴ Interim 2016 targets are not available; see MOGE 2015a for Zambia's national literacy targets for 2019.

⁵ TTL targets were set using the previous MOGE standards (see Ministry of Education, Science, Vocational Training, and Early Education 2014).

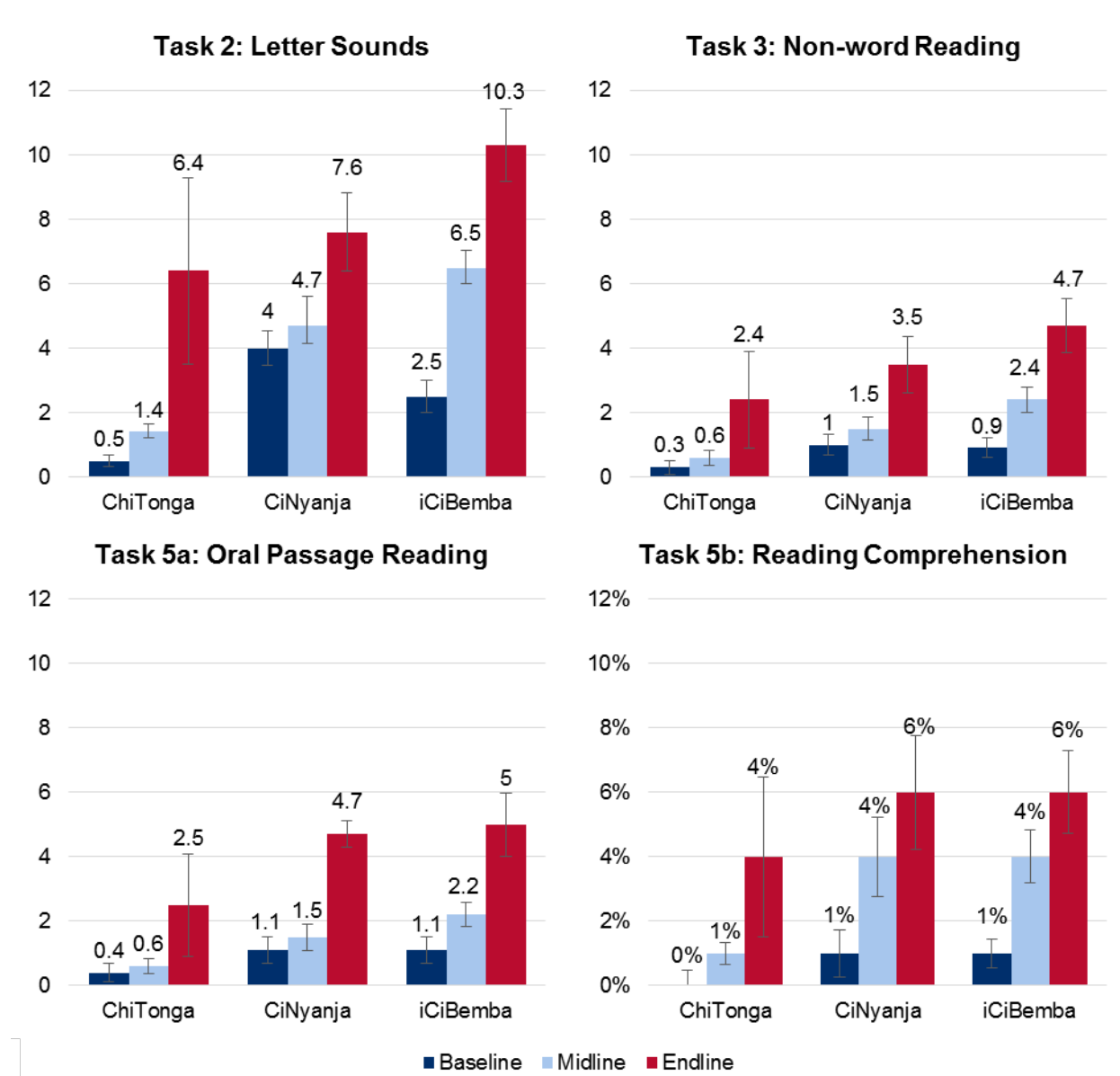
Exhibit 30: Distribution of All Learners across EGRA Performance Categories at Baseline, Midline, and Endline



[Exhibit 31](#) presents the mean scores for these four tasks by language and evaluation phase, unambiguously illustrating this upward trend in phonics and decoding skills (tasks 2 and 3), reading

fluency (task 5a), and reading comprehension (task 5b). CiNyanja and iCiBemba show significant improvements in all four tasks, while the increases in ChiTonga are significant in task 2 only.

Exhibit 31: Mean Scores by EGRA Task at Baseline, Midline, and Endline



2

Community school learners have shown significant and large improvements in phonics, decoding, and fluency reading skills since baseline, with some variance by EGRA task and province.

Exhibits 32, 33, and 34 present the distribution of learners across the four MOGE benchmarks for each evaluation phase at the provincial level for three EGRA tasks. These results show that significant improvements in phonics and decoding skills (tasks 2 and 3) and reading fluency (task 5a) have occurred in almost all provinces for most tasks, but also reveal the variation seen in reading results at the provincial level. This variation indicates that while trends are in the upward direction (as noted in Finding 1), provinces are not improving at the same rate.

The significant improvement in mean scores for all three tasks (letter sounds, non-word reading, and oral passage reading) in Central and Copperbelt provinces reveals unambiguous progress (see [Exhibits 50, 51, and 53](#) in Annex 1). Comparing baseline to endline, learners in Central Province are reading, on average, 8.9 more letters sounded per minute, 4.7 more correct non-words read per minute, and 5.2 more correct words read per minute. Similarly, in Copperbelt, learners are reading an average of 8.3 more letters sounded per minute, as well as 3.8 more correct non-words read and 3.7 more correct words read per minute. Although most provinces showed gains for these three tasks that are typically considered large, the size of the gains in Central and Copperbelt provinces stand out as particularly strong, with effect sizes greater than 1.1 for letter sounds, 0.7 for non-word reading, and 0.6 for oral passage reading. As [Exhibit 32](#) illustrates, the shift in letter sound knowledge means more readers are performing at the emergent benchmark than 4 years ago, when the majority could not sound a single letter and more than 9 out of every 10 learners could not read a single word ([Exhibit 34](#)).

As shown in [Exhibits 50, 51, and 53](#) in Annex 1, Lusaka and Eastern provinces showed significant improvement from baseline to endline in mean scores on two of three tasks: Eastern Province on letter sounds (5.7 more letters sounded per minute) and oral passage reading (4.1 more words correct per minute) and Lusaka Province on non-word reading (1.4 more correct non-words per minute) and oral passage reading (1.6 more correct words per minute). Effect sizes of these improvements were also large, although smaller than observed in Central and Copperbelt provinces. In Lusaka Province, however, the distribution of scores for non-word reading showed no significant change across the evaluation phases, with a decline of only 3 percentage points in the proportion of learners unable to sound out a single non-word. This indicates that the shift in mean score is likely due to a very few exceptional learners at endline, rather than a fundamental shift in most learners' ability. Eastern Province was close to the 5 percent significance threshold for demonstrating a significant improvement in mean oral passage reading (task 5a) scores, and there was a significant shift upward in the distribution of learners performing at pre-emergent and emergent reading levels. Conversely, scores for letter sound knowledge (task 2) showed a significant but negative difference in Lusaka Province, with substantially more learners scoring zero at midline and endline than at baseline, indicating a stagnation or decline in ability on this task.

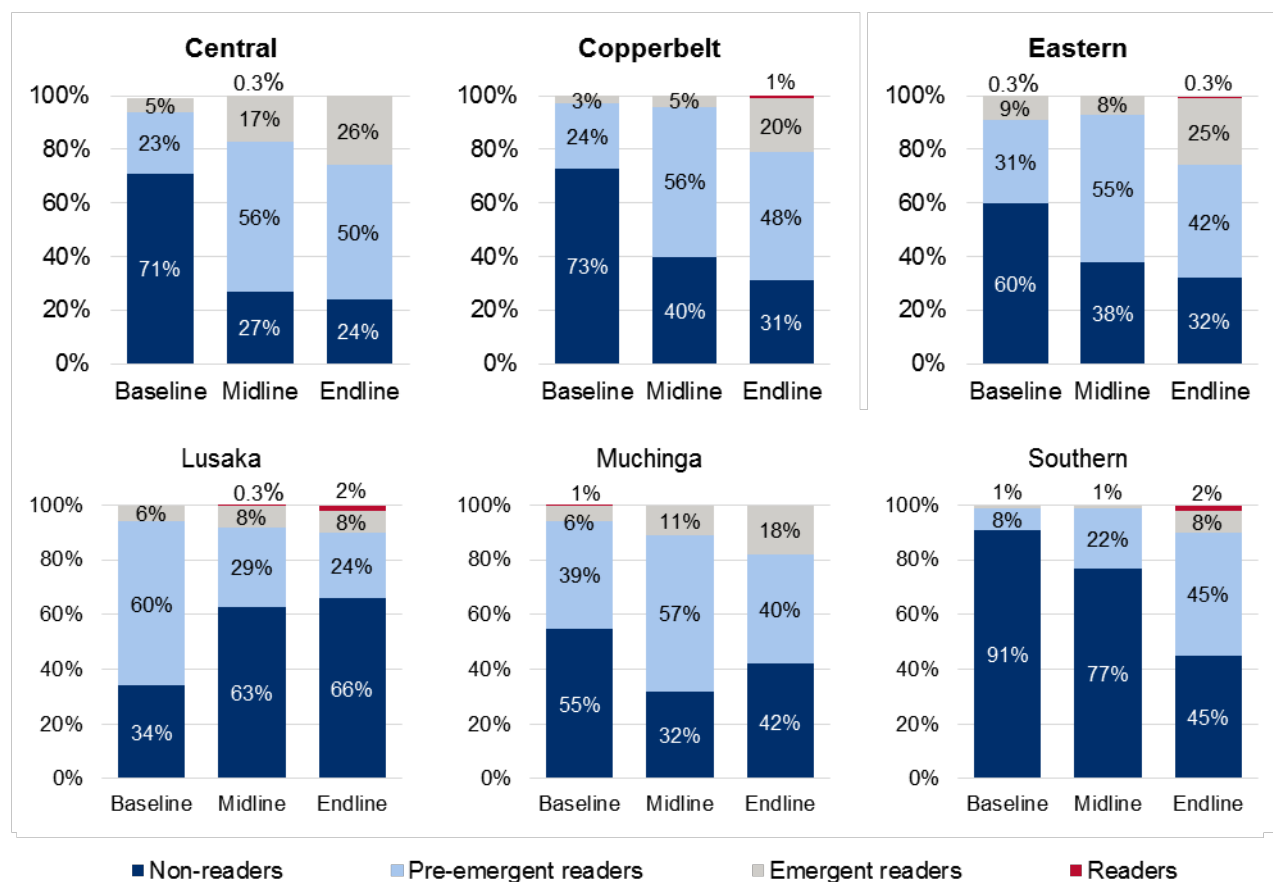


A qualitative shift is occurring, with the majority of learners now taking their first steps toward learning to read.

Muchinga and Southern provinces showed significant improvement in mean scores only for letter sound knowledge (task 2), with increases of 4.2 letters per minute in Muchinga and 5.9 letters per minute in Southern (see [Exhibit 50](#) in Annex 1). In Muchinga Province, there was similarly no change in the distribution of learners' scores across the four performance categories for non-word reading and oral passage reading. For Southern Province, the size of the improvement in letter sound knowledge was quite large. Southern Province is unique, in that it was a low performer at both prior measurement phases; 91 percent of learners could not sound a single letter at baseline, and performance on this task had barely moved at midline. Conversely, at endline, the majority (55 percent) of learners were able to sound at least one letter, closing some of the gap with other provinces.

For all provinces but Lusaka, a majority of learners scored zero at baseline on letter sounds; at endline, the majority were scoring above zero. For Central, Copperbelt, and Muchinga provinces at midline, sampled learners increasingly performed at the pre-emergent benchmark, and by endline a number of sampled learners achieved the emergent category. For letter sounds, four provinces now have double-digit proportions of learners performing at the emergent benchmark for letter sounds and nonsense words (the proportion is logically higher for letter sounds than nonsense words); for oral passage reading, the shift has been more from non-reader to pre-emergent, although growth has also been observed at the emergent benchmark (albeit slower than for the other two tasks).

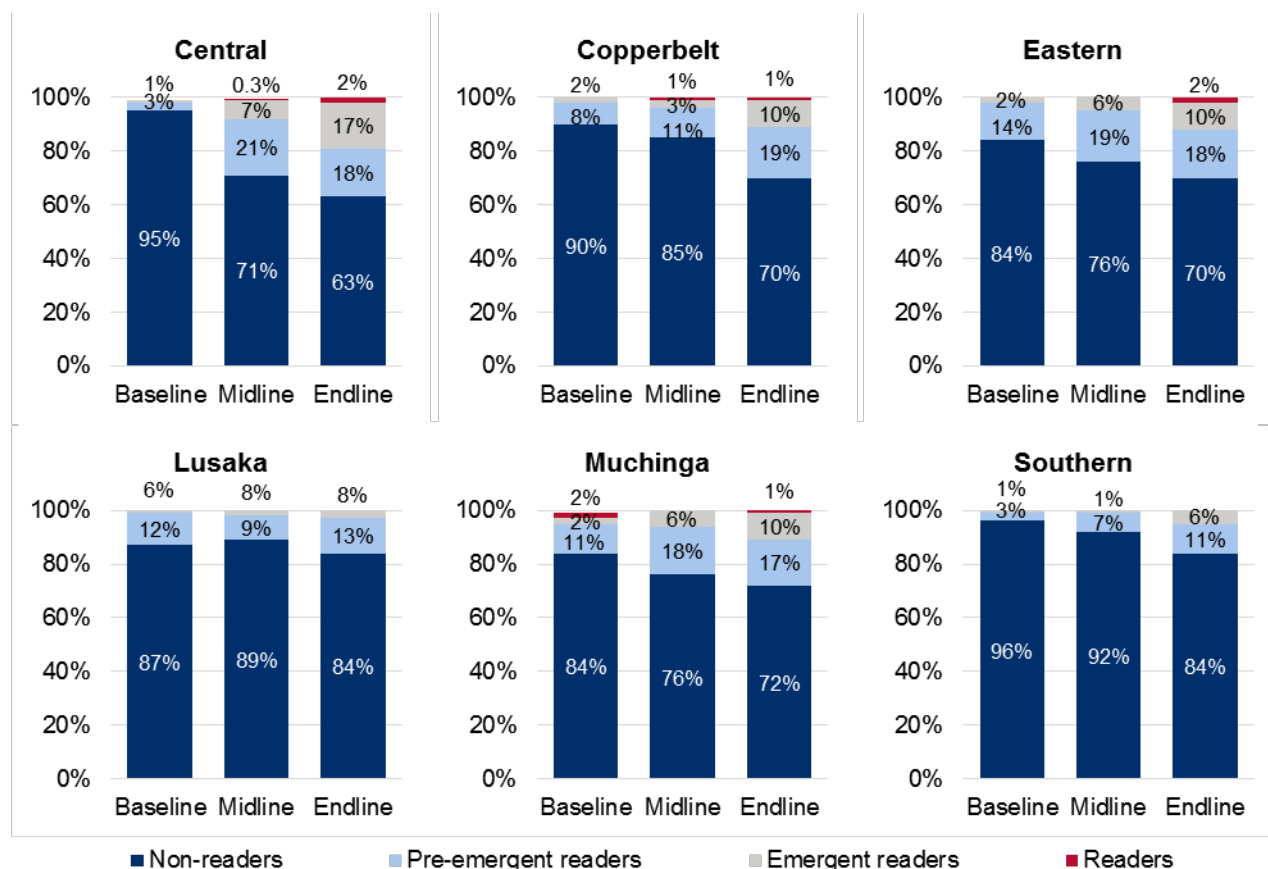
Exhibit 32: Distribution of Task 2 Scores (letter sound knowledge) by Province at Baseline, Midline, and Endline



Note: Differences across the three evaluation phases in the distribution of learners by these four categories is significant at the $p < 0.001$ level for all six provinces.

Data for this exhibit is available in [Exhibit 51](#).

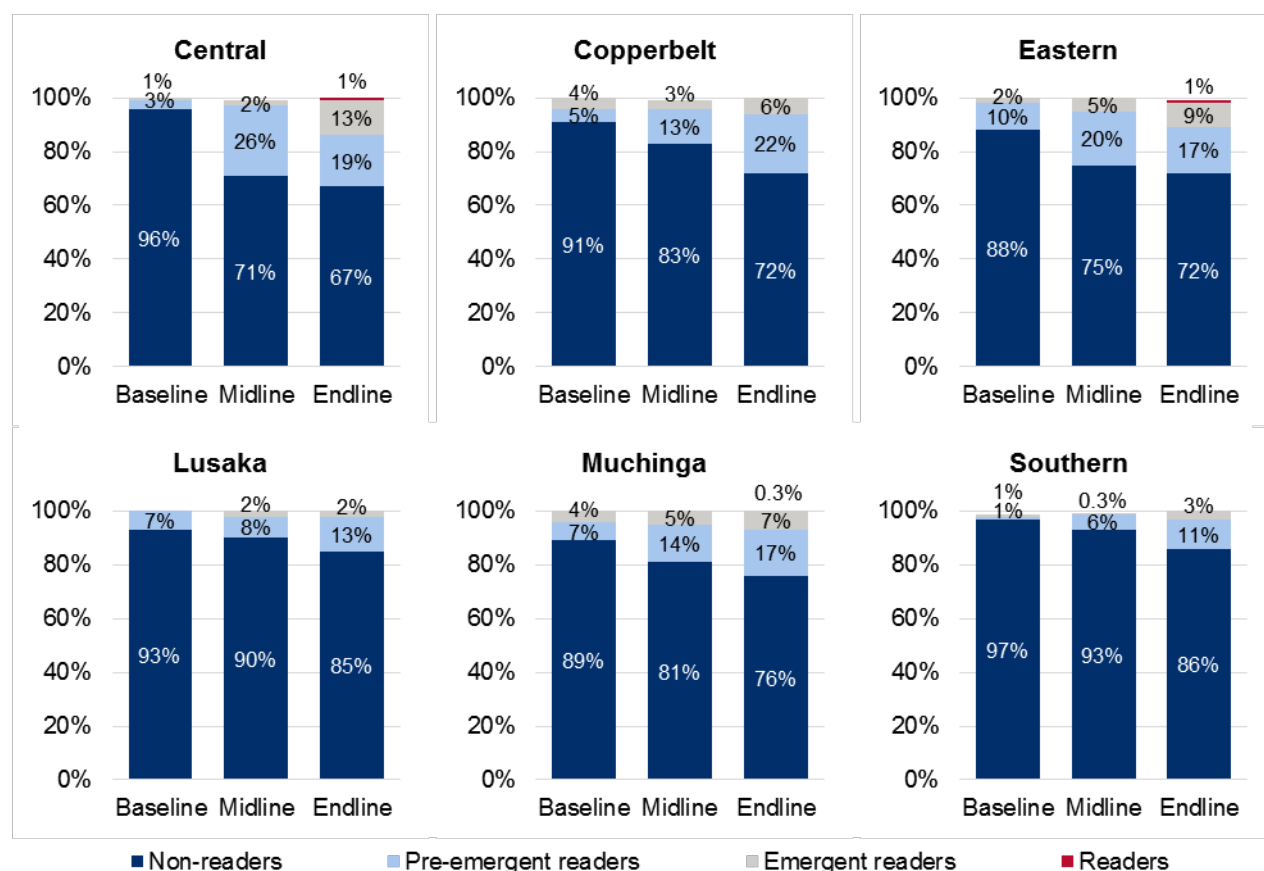
Exhibit 33: Distribution of Task 3 Scores (non-word reading) by Province at Baseline, Midline, and Endline



Note: Differences across the three evaluation phases in the distribution of learners by these four categories is significant at the $p < 0.001$ level for all provinces except Lusaka and Muchinga.

Data for this exhibit is available in [Exhibit 52](#).

Exhibit 34: Distribution of Task 5a Scores (oral passage reading) by Province at Baseline, Midline, and Endline



Note: Differences across the three evaluation phases in the distribution of learners by these four categories is significant at the $p < 0.001$ level for all six provinces except Lusaka and Muchinga. Lusaka is significant at the $p < 0.05$ level.

Data for this exhibit is available in [Exhibit 54](#).

3 Listening comprehension in the language of instruction (task 1) remains largely unchanged.

Oral language ability in the language of instruction, as measured by listening comprehension (task 1), has not shown the kinds of progress seen in phonics, decoding, and reading fluency from midline to endline (baseline data are not statistically comparable). While performance on task 1 is comparatively better than other tasks, it is substantially below the desired level (see [Exhibit 49](#) in Annex 1). At endline, learners answered 38–68 percent of questions correctly across provinces; this corresponds to correct answers on two or three of the five questions. The Zambia EGRA uses literal comprehension questions for this task, which is a lower-order skill than answering an inferential question.

Inability to answer questions correctly after being read a story orally indicates potential challenges of *linguistic* comprehension. This is important because oral language ability is a foundation skill for literacy. Children who are unable to understand the oral language are hampered in their ability to understand written language; they cannot read with comprehension. Because the listening comprehension tasks on the Zambia EGRA measure performance in the language of instruction, poor scores on this task may also mean that a child does not understand classroom instruction, which would impede learning more broadly.

Across the sample, approximately 20 percent of learners reported not speaking the language of instruction at home (see [Exhibit 24](#)). At endline, learners who did speak the language of instruction at home performed moderately better on listening comprehension, but the result is only significant at the 10 percent threshold (8 percentage points better; $p=0.062$). The data do not include any nuance about *how much* the language of instruction is spoken at home, potential dialectical differences in *how* the language is spoken at home or in the catchment area, or *how many languages* are spoken at home, factors that might help explain the impact of Zambia's linguistic diversity on learners' listening comprehension in the language of instruction. Similarly, the Zambia EGRA does not measure foundational linguistic or cognitive skills that underlie listening comprehension, such as vocabulary, grammar knowledge, working memory, or attention control, factors that may explain listening comprehension scores at the individual level.

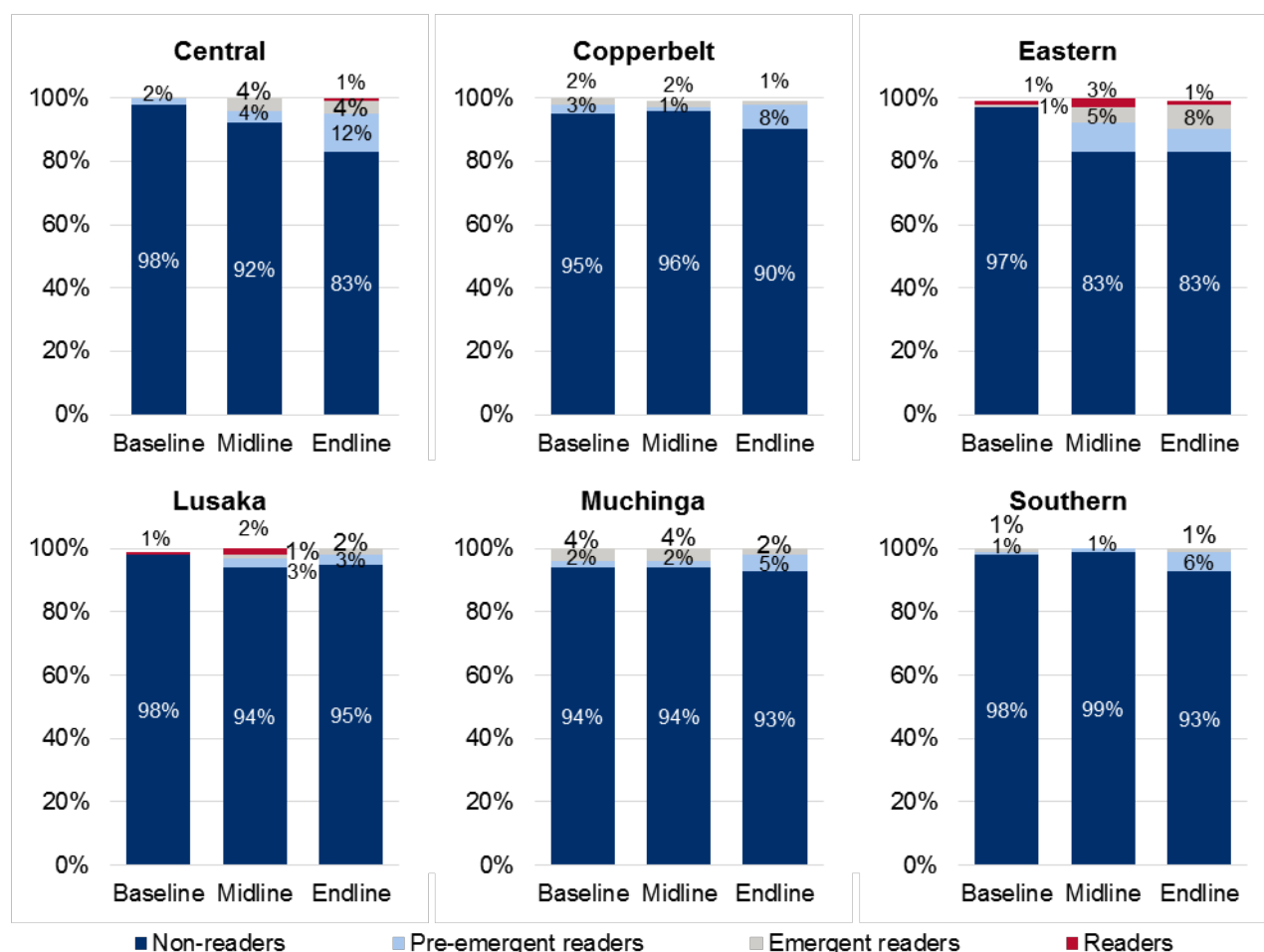
Scores on oral language ability are higher for the language of instruction than those for English listening comprehension (task 6), but similarly, there has been no significant change in English listening comprehension ability since midline. Mean scores remain low across provinces, ranging from a low of 4 percent (Muchinga) to a high of 40 percent (Copperbelt) for the number of correct responses out of five questions. This performance is expected, since learners are intended to begin oral English only in grade 2. TTL interventions have not focused on improving oral English in learners; the task is included in the assessment for its added contextual value. The differences in oral comprehension in the language of instruction and English indicate that, despite Zambia's linguistic diversity, learners seem to comprehend better in the languages of instruction than in English.

4

Central, Copperbelt, and Eastern provinces have all shown significant and large improvements in reading comprehension scores, and Lusaka and Southern provinces have shown significant upward shifts in the distribution of scores. There was no shift in scores in Muchinga Province.

Central, Copperbelt, and Eastern provinces have all shown significant and large improvements in reading comprehension mean scores, as measured by EGRA task 5b (see [Exhibit 55](#) in Annex 1). In Copperbelt Province, the improvement in mean scores appears to have been driven by a decrease in non-readers and commensurate increases in learners performing at the pre-emergent level, while Central and Eastern have also seen more learners at endline performing at the higher end of the distribution (at higher MOGE benchmarks), with learners shifting from the pre-emergent to the emergent category. As seen in [Exhibit 35](#), these upward shifts in the distributions across evaluation phases were also significant.

Exhibit 35: Distribution of Task 5b Scores (reading comprehension) by Province at Baseline, Midline, and Endline



Note: Differences across the three evaluation phases in the distribution of learners by these four categories is significant at the $p < 0.001$ level for Central, Copperbelt, and Eastern. Lusaka and Southern are significant at the $p < 0.05$ level. Data for this exhibit is available in [Exhibit 55](#).

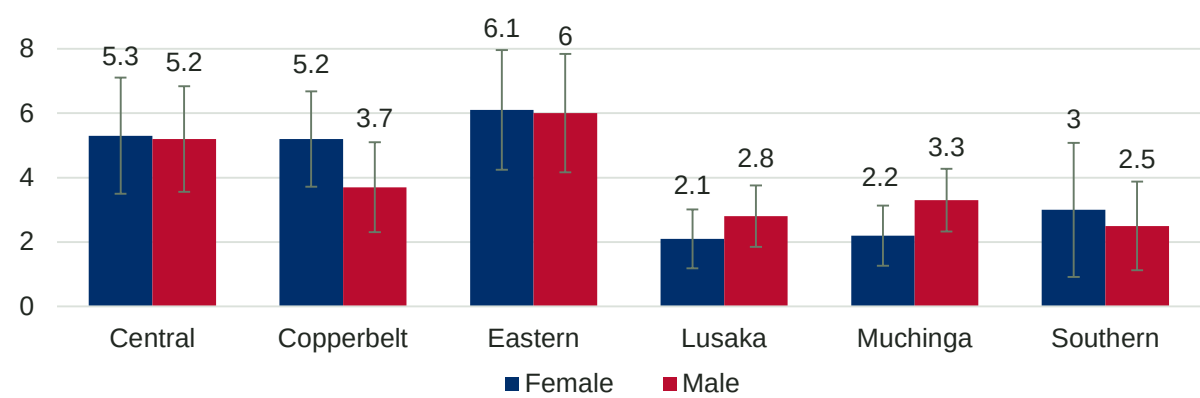
In Lusaka and Southern provinces, there was a significant decrease in the number of non-readers from 98 percent to 95 percent in Lusaka and from 98 percent to 93 percent in Southern, while there was no significant change in the mean score—a positive improvement, indicating that learners are making the qualitative shift from no reading comprehension ability to early skill development. In Lusaka, Muchinga, and Southern provinces, the low mean oral passage reading scores indicate that a majority of learners did not read a sufficient portion of the passage to have been asked the first comprehension question (see [Exhibit 54](#) in Annex 1). Thus, in these provinces, the lack of progress in reading comprehension is primarily indicative of a lack of fluent reading ability; it is unknown whether learners would be able to answer comprehension questions once they mastered this preceding task.

5 Across all three phases, parity in performance of male and female learners has been maintained, even as performance has improved over time.

The baseline found that, overall, male and female learners in community schools performed similarly in terms of reading outcomes. The improvements in reading scores observed at midline and endline (Findings 1 and 2) have occurred among male and female learners and, as a result, this parity has been

maintained. [Exhibit 36](#) presents the mean oral passage reading (task 5a) endline scores for both sexes by province, showing the similarity between male and female learners that reflects this parity in performance. Although reading ability remains low in absolute terms, gains have not been achieved at the expense of either sex, and community schools have shown a 4-year ability to avoid further marginalizing groups that may be at societal disadvantage.

Exhibit 36: Mean Scores for Task 5a (oral passage reading) at Endline by Province



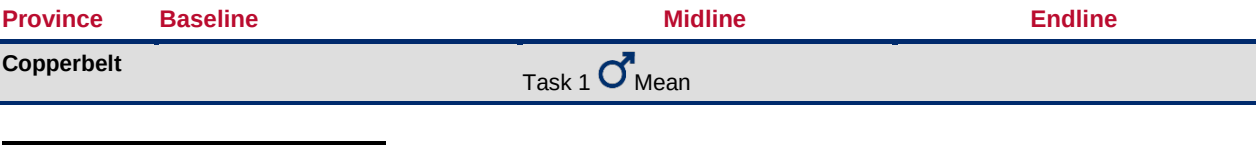
Data for this exhibit is available in [Exhibit 70](#).

In total, across all three phases, only 25 of 114 comparisons⁶ between male and female performance showed a significant difference. [Exhibit 37](#) consolidates these few significant differences on all EGRA tasks across three evaluation phases and six provinces (see [Exhibit 65](#) through [Exhibit 71](#) in [Annex 1](#) for full sex disaggregated results). While males are more likely to perform better than females in most of these few instances where performance differed significantly, there is no consistent pattern across evaluation phase, province, or EGRA task. Comparing means, only 22 percent of comparisons (25 out of 114) showed a significant difference between female and male learners: 7 at baseline, 12 at midline, and 6 at endline. In these 25 significant differences, 22 favored males and 3 favored females. Comparing the distribution of scores across the four reading performance categories, only 15 percent of comparisons (17 of these 114) showed a significant difference between female and male learners: 5 at baseline, 8 at midline, and 4 at endline; 13 of these differences favored males, 2 favored females, and 2 were ambiguous. Ultimately, across the full range of these comparisons, there is no consistent pattern to this handful of significant differences between sexes or among evaluation phases, provinces, and EGRA tasks. This affirms that for the past 5 years, similarity in reading scores between male and female learners has been the norm, despite a few instances of differences.



Children are achieving similar outcomes in TTL-supported community schools, regardless of sex.

Exhibit 37: Significant Differences by Sex, Province and EGRA Task



⁶ The 114 comparisons between male and female performance include three phases multiplied by seven tasks and six provinces, minus the two EGRA tasks not included at baseline.

Province	Baseline	Midline	Endline
		Task 2  Mean Task 3  Mean Task 5a  Mean and distribution Task 5b  Mean	Task 2  Mean Task 4  Mean
Central		Task 1  Distribution Task 3  Mean Task 5a  Mean Task 5b  Mean and distribution	Task 1  Mean,  Distribution
Eastern	Task 2  Mean Task 3  Mean and distribution Task 5a  Mean and distribution	Task 6  Mean and distribution	
Lusaka	Task 3  Distribution Task 5a  Distribution		Task 1  Mean
Muchinga	Task 1  Mean and distribution Task 3  Mean Task 5a  Mean Task 5b  Mean	Task 4  Mean and distribution Task 6  Mean and distribution	Task 2  Distribution Task 6 Distribution (neither sex favored)
Southern		Task 2  Distribution Task 4  Mean	Task 1  Mean Task 3 Distribution (neither sex favored) Task 6  Mean
TOTAL	7 Mean, 5 Distribution	12 Mean, 7 Distribution	6 Mean, 4 Distribution

Note: Symbol indicates the sex that performs better by task at significance of $p < 0.05$.

6

Community school learners exhibit great variation in reading ability, both across and within schools.

Community schools and their learner populations are not homogenous. At midline and endline, reading outcomes fluctuated widely from one school to the next. At some schools, almost all of the learners sampled were non-readers, while other schools had comparatively exceptional learning outcomes. In

other schools, the data show learners performing at the high and low ends, as well as points in between. Across schools, two distribution patterns among learners are evident: (1) scores clustering tightly, either at the high or low end of the performance spectrum, indicating that the school achieved fairly consistent learning outcomes across learners; and (2) a wide range of reading scores within the same school, suggesting that some learners were doing well and others were being left behind.

[Exhibit 38](#) and [Exhibit 39](#) illustrate these patterns for letter sound knowledge (task 2); the wider variation in scores for this task makes it easier to view the trend, although the trend is also evident for oral passage reading scores (task 5a). In the exhibits, each vertical band represents a single school; schools are sorted from left to right on the median EGRA score, in ascending order, meaning schools to the right have higher medians. The blue box for each school represents the values between the 25th and 75th percentiles. The thin vertical lines extending in both directions from the box represent the farthest points that are not outliers (“adjacent values”), and the dots represent outlier learner scores. Some schools demonstrate a shorter vertical band, meaning more consistent scores across their learners, while others appear more stretched, indicating a wider distribution of scores within the school. The dashed red horizontal line represents the average score across all learners at that phase, reflecting schools’ wide variation around the overall mean (which is low), and the extent to which many schools themselves are outliers.



Learner performance at community schools is widely variable, both within and across schools.

The data are presented for midline ([Exhibit 38](#)) and for endline ([Exhibit 39](#)), showing that the overall wide variation has persisted (and possibly even grown) across—and in some cases within—schools. Although 34 percent of endline learners sampled attended a school where the median letter sound knowledge score was zero, this was an improvement over midline (when about 45 percent of learners attended a school with median score of zero). Nonetheless, this indicates that a large portion of learners are attending schools where little learning has been happening, though this does not imply that the schools are the cause. Conversely, at endline, about 32 percent of sampled learners attended a school where the median was greater than 8 letters per minute, meaning almost an equal proportion of learners at endline attended model schools, where more than half of the children outperformed the national community school average.

Exhibit 38: Letter Sound Learner Performance by School, Midline 2014
(*n=1,663 learners and 102 schools*)

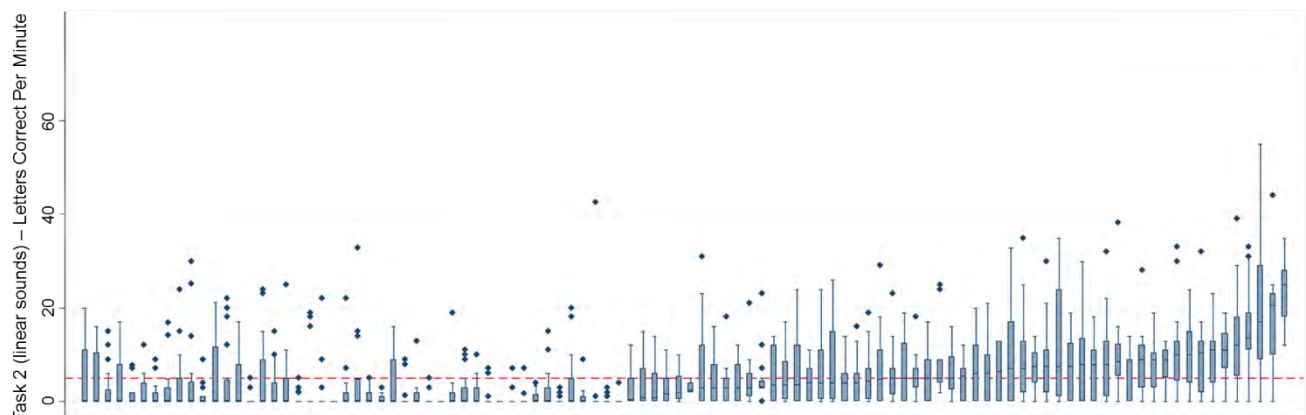
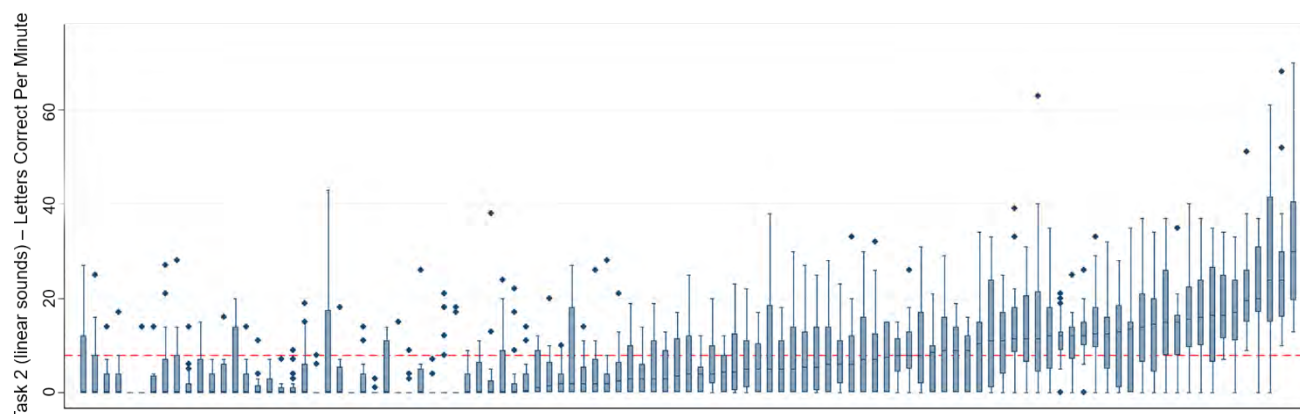


Exhibit 39: Letter Sound Learner Performance by School, Endline 2016

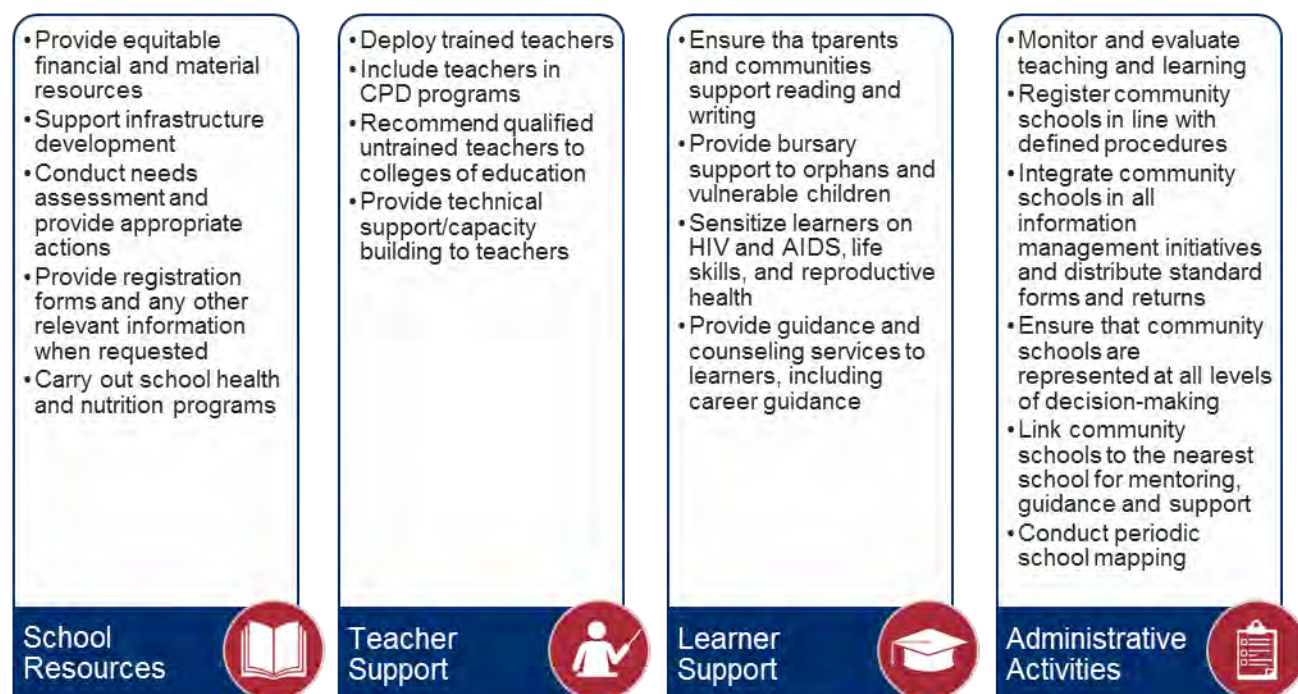
(*n=1,787 learners and 105 schools*)



5.2. MOGE SUPPORT TO COMMUNITY SCHOOLS

Zambia's 1996 *Educating Our Future* policy mandated the MOGE to provide “quality education” to all eligible children and established the country's overall framework for education delivery. Recognizing the important role that communities play in the provision of education, the MOGE developed the OGCS in 2007 to support implementation of the policy as it pertains to the management and support of community schools. The 2011 Education Act officially recognized the role of community schools in Zambia's education system, and in 2014, with TTL support, the MOGE revised the OGCS to reflect the new education act and further improve the policy framework governing community schools. The 2007 and 2014 OGCSs both provide a framework for supporting children in community schools. They are intended to: (1) guide MOGE officials and other stakeholders on their roles and responsibilities with regard to management and coordination of community schools; and (2) commit the Zambian government to infrastructure provision, teacher training, government teacher deployment, and equitable resource allocation to community schools. All evaluation phases collected data based on the MOGE roles and responsibilities, stipulated in the OGCS and presented in [Exhibit 40](#).

Exhibit 40: MOGE Roles and Responsibilities in the OGCS



Source: MOGE (2016b)

This section presents findings from three data sources: the community school head teacher questionnaire, the post-observation teacher interview, and the self-administered survey questionnaire to MOGE District Education Board Secretary offices. All three tools focused on seven⁷ MOGE responsibilities to community schools:

- Direct financial support
- Teaching and learning materials (e.g., learners' books, teachers' guides)
- Free basic materials (e.g., chalk, exercise books, pencils)
- CPD (training for teachers and head teachers)
- Monitoring visits
- Infrastructure support, including building materials
- Government teachers deployed to community schools.

Baseline, midline, and endline data presented below represent the percentage of schools reported by the head teacher or the MOGE to have received that type of support in the past year.

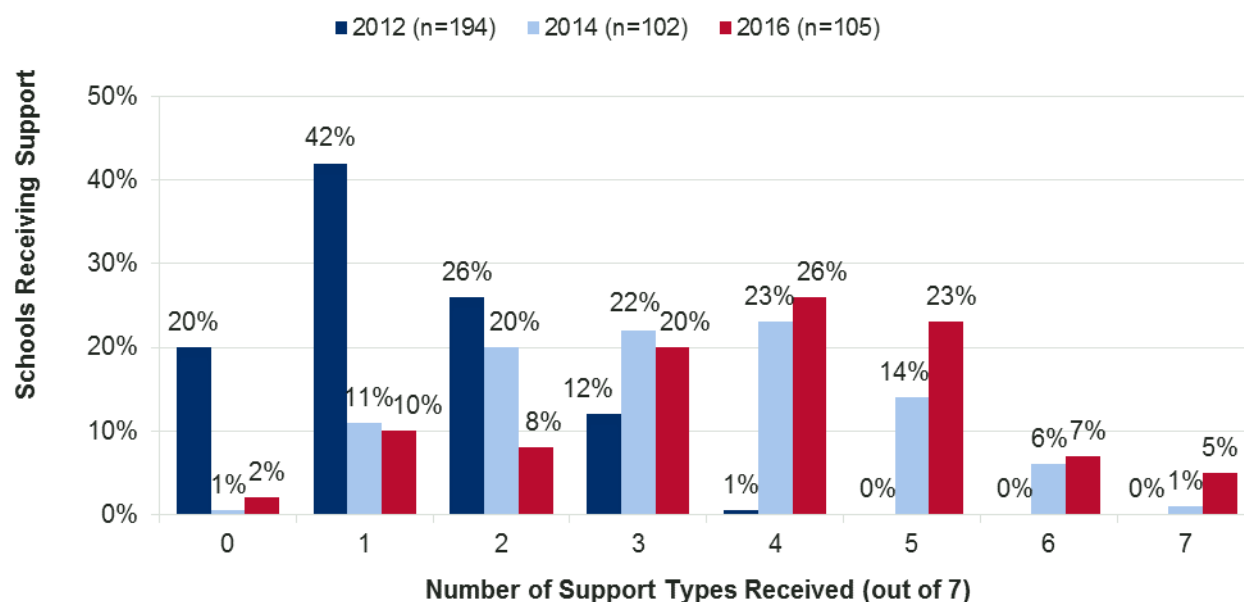
7 Head teachers report receiving significantly more types of MOGE support from 2012–2014 and from 2014–2016.

Sampled community schools have received a significant increase in median number of types of MOGE support over time ($p < 0.001$). The median number of types of MOGE support received increased

⁷ At baseline and midline, the District Education Board Secretary survey did not ask about the last two types of support, infrastructure and provision of government teachers.

significantly from one at baseline to three at midline ($p<0.001$) and further still at endline, to four out of the possible seven (midline to endline, $p=0.007$). When teachers were asked how many forms of support they received from the MOGE in 2012, 62 percent responded zero or one, and 99 percent indicated three forms or fewer. By contrast, at endline, a strong majority (61 percent) indicated that their schools received four or more forms of support ([Exhibit 41](#)).

Exhibit 41: Number of MOGE Support Types Received over Time, as Reported by Head Teachers



Presence of a government-seconded teacher and location of the school are both associated with the number of MOGE support types received. At midline and endline, schools with a government-seconded head teacher were significantly more likely to have received one or more additional forms of support ($p<0.001$). At endline, rural schools received more types of support than urban schools ($p<0.001$).

8 | Significantly more community schools are receiving the majority of specific types of MOGE support in 2016 than in 2012.

According to head teachers, the proportion of schools receiving the following six types of support has increased significantly since 2012, as presented in [Exhibit 42](#):

- Teaching and learning materials (75 percentage-point increase, $p<0.001$)
- Free basic materials (57 percentage-point increase, $p<0.001$)
- Teacher training (31 percentage-point increase, $p<0.001$)
- Monitoring (46 percentage-point increase, $p<0.001$)
- Deployment of government teachers (39 percentage-point increase, $p<0.001$)
- Infrastructure (6 percentage-point increase, $p=0.027$).

Five of these areas of MOGE support (teaching and learning materials, free basic materials, teacher training, monitoring, and deployed government teachers) increased significantly from 2012–2014, and three areas (teaching and learning materials, deployed government teachers,

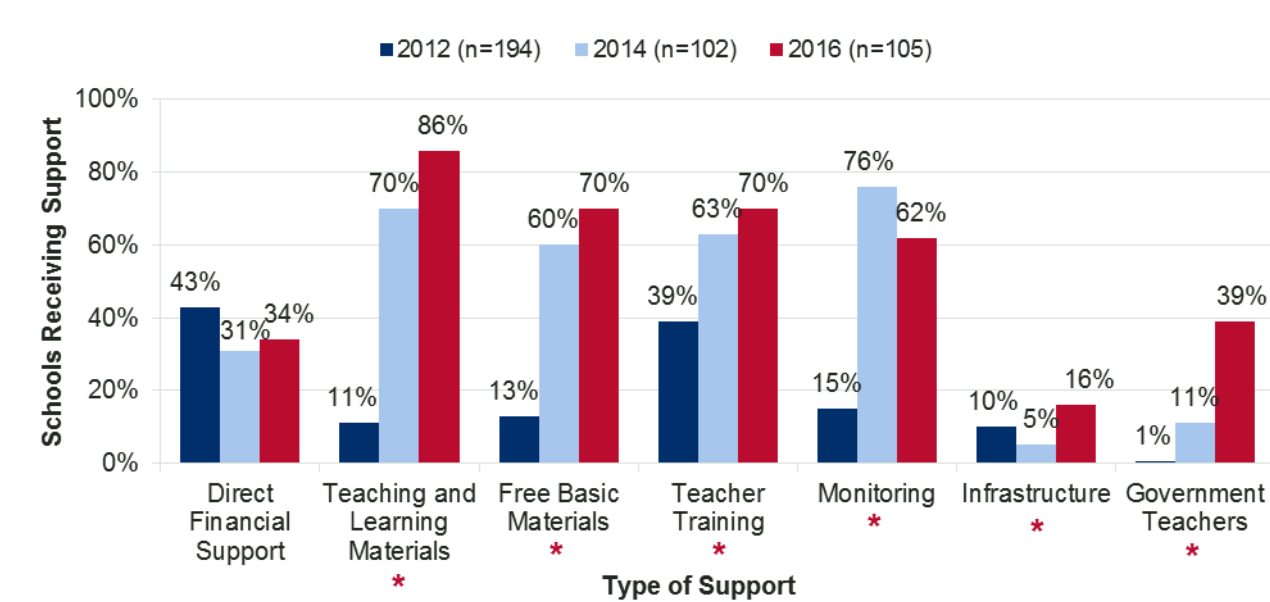


The MOGE is reported as the primary provider of non-monetary support to community schools.

and infrastructure) continued this trend in significant increases from 2014–2016. The largest increase from midline to endline was the placement of government teachers at community schools (28 percentage-point increase, $p=0.023$), with 39 percent of community schools sampled at endline having received a government teacher. Government teachers are an increasing portion of the community school workforce, comprising 38 percent of all community school teachers at endline, compared to 26 percent at midline.

Although the proportion of sampled community schools that was monitored at least once by the MOGE in the previous year increased significantly from 2012–2014, there was a significant decrease (14 percentage points, $p=0.023$) in the number of schools monitored by MOGE from 2014–2016. Nonetheless, the proportion of community schools that had been monitored by the MOGE remained significantly higher at endline than at baseline.

Exhibit 42: Comparison of Specific Types of MOGE Support Provided in 2012, 2014, and 2016, as Reported by Head Teachers



* Denotes a difference in the proportion of schools receiving support that is significant at 5 percent.

Rural community schools were significantly more likely than urban schools to receive a government teacher (33 percentage points, $p=0.004$) and to receive training ($p<0.001$), free basic materials ($p<0.001$), and infrastructure ($p=0.017$). However, urban and rural schools receive similar amounts of monitoring, teaching and learning materials, and direct financial support. Schools without a government-seconded teacher received increased monitoring ($p=0.007$). While a significant difference on whether a school had received MOGE training or not existed across provinces at baseline ($p<0.001$), no significant difference was found at endline; this decrease in variation across provinces is a positive sign of increasing national harmonization of support to community schools.

The MOGE is the primary provider of non-monetary support, with the majority of schools reporting that they received more from the MOGE than from other sources. On the whole, however, head teachers still reported that community schools were under-resourced, despite the large and significant increase in the number of sampled schools that the MOGE is supporting. Across the schools sampled at endline, schools averaged one classroom per 71 learners. Access to toilets is adequate for teachers (3:1 for both male and female teachers), but learner access is lower (46:1 for male learners and 47:1 for female learners). On average, head teachers reported just one office where school materials could be stored securely, and 63 percent of schools did not have safe rooms to hold exams. Head teachers reported

inadequate benches (70 percent), teacher desks and chairs (80 percent), chalkboards (43 percent), basic classroom materials (52 percent), literacy textbooks (78 percent), other teaching and learning materials (66 percent), and water supply (53 percent). There was no significant variation across provinces or urban and rural schools.

9

The number of head teachers who report receiving a monitoring visit from a MOGE official has increased significantly since baseline and MOGE officials are observing more literacy classes, as compared to midline.

As [Exhibit 42](#) shows, monitoring by MOGE officials showed a 46 percentage-point increase from baseline to endline. At endline, head teachers reported that their schools had received a mean of 1.4 monitoring visits (median=1) in the previous year. Significance could not be tested across provinces, but the variation is interesting and warrants additional research: Schools in Lusaka Province received 2.7 MOGE monitoring visits; Central received 1.9; Muchinga 1; and Copperbelt, Eastern, and Southern 0.9 each. Zonal-level MOGE officials were most likely to have visited a school at least 1.0 time in the previous year, followed by district MOGE officials (0.3 times).

Although overall monitoring by MOGE officials increased significantly from 2012–2016, the percentage of schools receiving this support has decreased by 14 percentage points since midline. At endline, however, MOGE monitoring visits were much more likely to include classroom observation than at midline. During the most recent visit by a MOGE official, community school head teachers reported that the most common activity was observing a literacy class (58 percent), followed by providing feedback to the teacher on the lesson observed (54 percent) and recording school details such as enrollment levels and number of teachers (51 percent).

Sixty-one percent of teachers whose class was observed at endline reported that a MOGE official had observed their class in the last year, compared to 28 percent at midline ($p<0.001$). Among teachers reporting that MOGE officials had observed their classes, 27 percent said the MOGE official had reviewed lesson plans and 21 percent said the official had reviewed learner assessments. Fewer teachers reported that the MOGE official had modeled or demonstrated how to conduct a literacy lesson (12 percent), provided ideas regarding how to develop or use teaching and learning materials (12 percent), or assessed learners in literacy and provided specific feedback on their performance (12 percent). Only 12 percent reported that the MOGE official did nothing after observing the class.

The top five types of feedback provided by MOGE officials after observing a literacy class were (1) break words into sounds and syllables to help learners read difficult words (19 percent); (2) show learners how to read left to right (17 percent); (3) read stories to the learners and ask questions based on the story (17 percent); (4) give learners a chance to read (15 percent); and (5) review new vocabulary with children (14 percent).

10

When community schools report receiving direct MOGE funding, this support constitutes the bulk of the school's annual operating budget.

Based on head teacher data, the percentage of schools receiving a grant remained stable from midline (30 percent) to endline (34 percent), as did the average grant amounts. Sampled endline schools receiving grants got an average grant of 5,214 kwacha (\$520) per year, while sampled midline schools averaged 5,303 kwacha. At endline, schools in Lusaka Province were more likely to receive grants (30 percent), followed by schools in Southern Province (28 percent). The 2016 District Education Board Secretary survey reported giving 3,620 kwacha (\$360) per school, but to almost twice as many schools as those that reported receiving funds. Across baseline, midline, and endline, district offices reported distributing more direct financial support to community schools than head teachers reported receiving from the MOGE. This may be explained by how this type of support is understood; qualitative data

indicate that District Education Board Secretaries and some head teachers often use the term “grant” in a broad sense to refer to any kind of MOGE support, including in-kind as well as monetary grants.

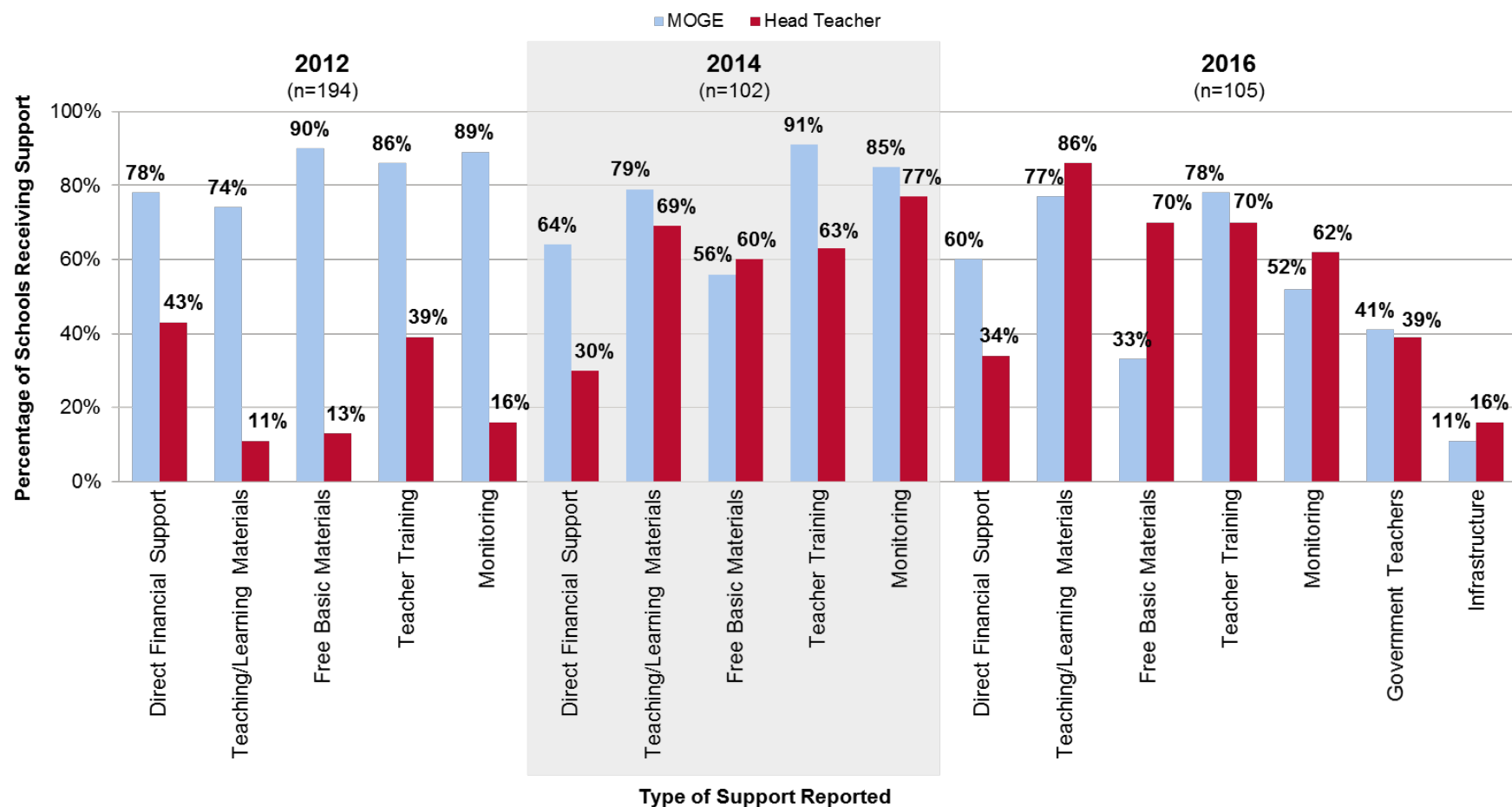
Only 48 percent of endline schools had an annual budget; the remaining 52 percent may have had financial resources, but did not have a school budget based on anticipated income and expenditures. For schools with an annual budget, MOGE financial contributions made up the majority (72 percent). Other sources of contributions to community school budgets were primarily from parents and community members (50 percent), school fees (36 percent), and income-generating activities (20 percent), which indicates community participation in and ownership of community schools. Only 18 percent of schools with an annual operating budget and 9 percent of sampled schools overall received funding from a nongovernmental entity. Most contributions to community schools from non-MOGE sources were teaching and learning materials (64 percent), teacher training (43 percent), and infrastructure (41 percent).



District Education Board Secretaries’ and head teachers’ perceptions of MOGE support to community schools have continued to converge over time.

MOGE district offices reported providing a high level of support to community schools at baseline, which differed greatly from the amount reported by head teachers. However, these perceptions have become more aligned over time ([Exhibit 43](#)). At endline, reports were the most aligned regarding seconded government teachers, which head teachers reported at 39 percent, compared to 41 percent by MOGE. This was followed by CPD provided to community school teachers, which was reported by 70 percent of head teachers and 78 percent of district offices, and teaching and learning materials, reported by 86 percent of head teachers and 77 percent of District Education Board Secretaries.

Exhibit 43: Community Schools Receiving MOGE Support in 2012, 2014, and 2016, as Reported by MOGE and Head Teachers



Note: The 2012 data from head teachers were collected at baseline. The 2012 data from the MOGE were collected retrospectively at midline.

Some differences in the reported support could be explained by the MOGE's reporting data for 2015 and 2016, whereas head teachers reported for 2016 alone. District Education Board Secretaries were often very specific in recalling the types of free basic materials, for example, and how they were distributed over both years, indicating more documentation at district level. This documentation of support to community schools seems to have increased since baseline, a further indication of increased MOGE responsibility toward community schools and their integration into Zambia's education system at the district level.

PCSCs also received MOGE support with head teachers reporting that 30 percent received assistance in developing school improvement plans and 12 percent received training. However, MOGE district offices reported providing training to 30 percent of PCSCs in 2016.

5.3. LITERACY INSTRUCTION IN COMMUNITY SCHOOLS

The 2013 *Zambia National Literacy Framework* is the guiding document for literacy instruction, introducing familiar language instruction for grades 1–4 and outlining five key early reading skills that form the basis of daily reading instruction in effective literacy programs ([Exhibit 44](#)): phonemic awareness, phonics, oral reading fluency, vocabulary, and comprehension (MOGE 2013a). Through technical assistance and partnership with the MOGE, TTL worked with education officers at all levels to develop strategies to implement MOGE reading instructional approaches at community schools. TTL supported teacher training on each of these five skills. In addition to pedagogical practice, training included classroom management tools to organize instruction and assess learners on an ongoing basis. Improving literacy instruction through these activities constituted a core strategy of TTL and the Primary Literacy Program's efforts to improve learning outcomes.

Exhibit 44: Zambia National Literacy Framework Key Literacy Instruction Skills

Skill	Description
Phonemic awareness ⁸	Ability to “hear” sounds and manipulate them orally (e.g., put sounds together, break words apart into sounds, and identify rhyming words, likenesses, and differences in spoken words)
Phonics ⁹	Ability to associate written letters (symbols) with their corresponding sounds
Oral reading fluency	Ability to read text with accuracy, speed, and expression
Vocabulary	Ability to understand the meaning of words and use them orally and in writing
Comprehension	Ability to understand the meaning of what is read or heard

Based on midline evaluation findings that showed inadequate distribution of instructional time across the five skills, in its last 2 years TTL focused its teacher training modules and supporting materials on the whole lesson and means of incorporating literacy skills through five steps: (1) discuss a topic or do a picture walk; (2) model fluent reading; (3) give learners the chance to read; (4) guide learners when they make errors and pause to ask comprehension questions; and (5) allow learners to use the

⁸ This skill is commonly referred to as phonological awareness, with phonemic awareness forming a sub-skill that captures the ability to hear the smallest units of sound that can distinguish words in a language. Phonemic awareness and phonological awareness are often used interchangeably (RTI International 2015).

⁹ This definition also encompasses the competences commonly referred to as the alphabetic principle and decoding. Phonics is also commonly known as the instructional method that teaches literacy based on these associations (ibid.).

information they read by writing, drawing, or answering questions. This training aligned with capacity building designed to empower zonal offices with the skills to provide appropriate and targeted feedback on the five key literacy areas after observing a literacy lesson.

This section presents findings based on three data sources: classroom observation, teacher interviews following classroom observation, and head teacher questionnaires. The endline classroom observation measured 26 distinct criteria of high-quality literacy instruction (20 of which are comparable to midline), with criteria grouped into five domains, reflecting the five key components of literacy instruction from the *Zambia National Literacy Framework*, plus a sixth domain to capture writing.

[Annex 2](#) provides detailed classroom observation results for midline and endline.

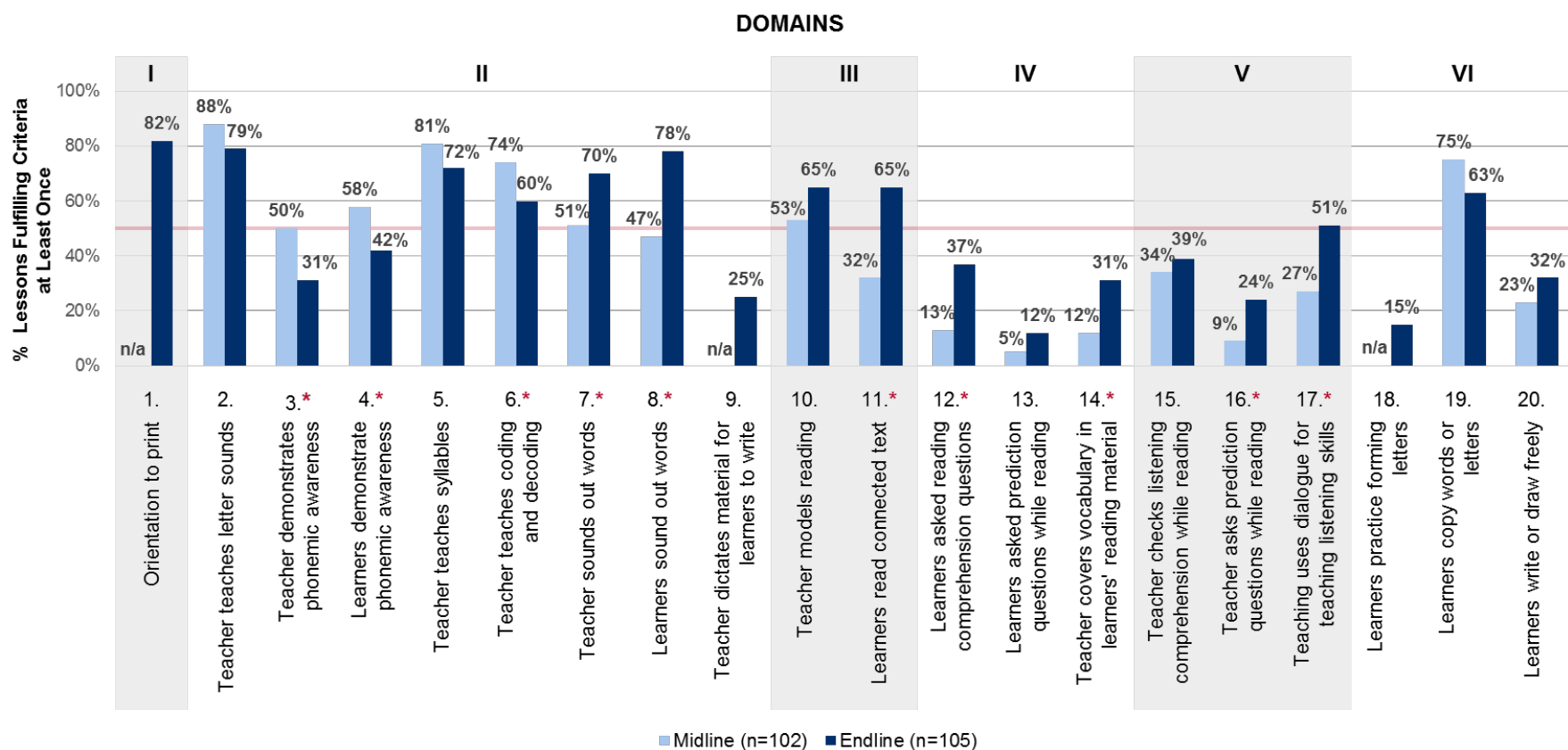
12

Observed lessons fulfilled criteria related to higher-order literacy skills, such as reading connected text and comprehension, significantly more at endline than at midline.

At midline, most observed lessons included, at least once, activities related to phonemic awareness and phonics (domain II), but fewer than half of the lessons spent any time on activities related to reading comprehension (domain IV) and oral language development (domain V), as illustrated in [Exhibit 45](#) (see [Exhibit 72](#) in Annex 2 for full results). While 53 percent of midline lessons included teachers modeling fluent reading (criterion 10), only 32 percent offered learners the opportunity to do so (criterion 11). This means a minority of the midline lessons observed included any activity related to criteria 11–17, illustrating a misbalance of instructional focus in favor of lower-order literacy competencies at the expense of higher-order ones.

At endline, conversely significantly more lessons fulfilled these same criteria at least once, with the exception of criterion 15 (teacher checks listening comprehension while reading or telling a story to learners), which did not show a significant change from midline. Additionally, significantly more lessons also featured activities related to sounding out words (fulfilling criteria 7 and 8) at least once at endline compared to midline. However, a majority of lessons still did not fulfill the individual criteria within the reading and listening comprehension domains at least once. Nevertheless, the significant increase from midline indicates that many lessons have, in some measure, started to become more balanced, with teachers increasingly incorporating at least some time focused on higher-order reading skills. Fewer lessons at endline focused exclusively on teaching individual letter sounds or basic decoding skills. This is a substantial improvement in lesson quality in just 2 years.

Exhibit 45: Percent of Lessons that Fulfilled Criteria at least Once – Comparison of Midline to Endline



* Denotes significance at 5 percent.

Domains: I-Orientation to Print; II-Phonemic Awareness and Phonics; III-Passage/Story Reading; IV-Reading Comprehension; V-Oral Language; VI-Writing

Similar patterns are observed when examining the average number of intervals each criterion was fulfilled per lesson (a proxy for the amount of lesson time devoted to that criterion), as seen in [Exhibit 46](#) (see [Exhibit 75](#) in Annex 2 for full results). At midline, observed lessons devoted substantially more instructional time to teaching letter sounds (criterion 2), syllables (criterion 5), coding and decoding (criterion 6), and having learners copy words or letters (criterion 19). On average, each lesson fulfilled these criteria four or more times over the course of a 20-interval (1-hour) lesson. (Criteria are not mutually exclusive; activities can fulfill multiple criteria simultaneously.) [Exhibit 46](#) indicates that although lessons observed at endline still devoted substantial instructional time to those tasks primarily falling in domain II (phonemic awareness and phonic), time on those tasks decreased significantly to range from 2–4 intervals at endline. There were also significant decreases from midline to endline on several other tasks in the phonemic awareness and phonics domain, including criterion 2 (teacher teaches letter sounds, 2.18 decrease, $p<0.001$), criterion 3 (teacher demonstrates phonemic awareness, 1.45 decrease, $p<0.001$), criterion 4 (learners demonstrate phonemic awareness, 1.5 decrease, $p<0.001$), criterion 5 (teacher teaches syllables, 1.86 decrease, $p=0.002$), and criterion 6 (teacher teaches coding and decoding, 1.32 decrease, $p=0.012$).

The decrease in instructional time spent on lower-order tasks allowed for more time for other activities, as seen by the significant increase at endline in the average number of intervals during which lessons fulfilled 10 of the 17 criteria (criteria 7–17, all of which increased significantly at the 5 percent level, except criterion 10, with an increase significant only at the 10 percent level). Particularly notable is the increased instructional time for the higher-order literacy skills of reading fluency (domain III) and comprehension (domains IV and V).

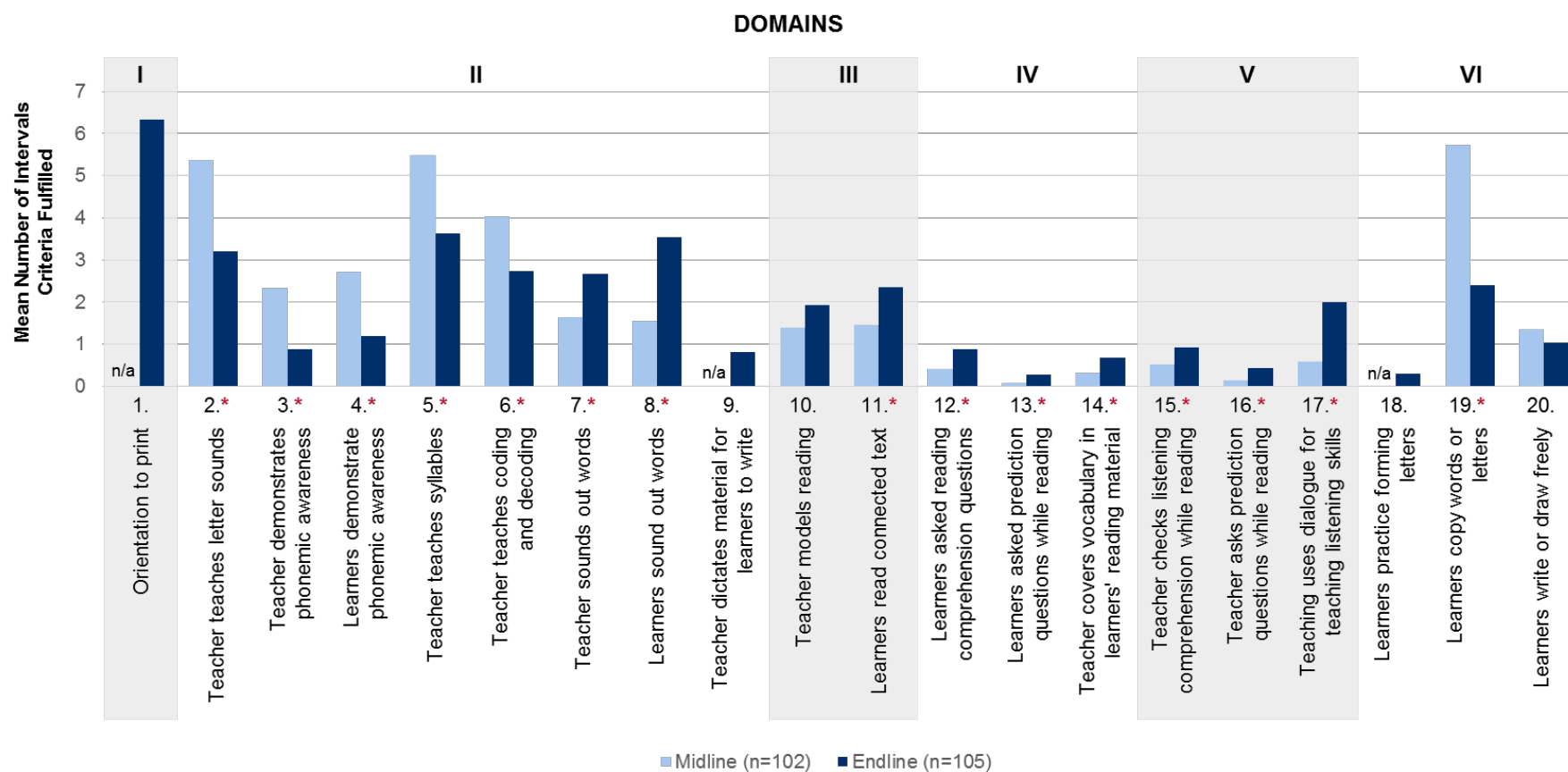


Community school teachers are beginning to incorporate higher-order reading skills into their literacy lessons.

These increases indicated that teachers were incorporating important changes into their literacy lessons; in absolute terms, however, the amount of instructional time devoted to these higher-order skills remained low. At endline, for example lessons fulfilled criterion 11 (learners read connected text) for an average of 2.4 intervals; because intervals cover 3-minute periods, this indicates that, on average, no more than 7.5 minutes were spent on learners reading connected text aloud during a lesson. This time allocation leaves minimal opportunity for the majority of learners to participate in these activities in the community school context, where the average teacher-to-learner ratio is 1:60. The amount of time devoted to the criteria related to listening and reading comprehension is even less, with all six criteria in these two domains fulfilled for less than 1 interval on average per lesson, indicating that teachers gave a maximum 3 minutes, on average, to these topics.

At endline, the average length of observed literacy lessons was 51 minutes, which is similar to the average observed lesson length at midline (49.6 minutes). The endline observed lesson length is a significant increase from baseline (37.7 minutes, $p<0.001$) and nears the MOGE standard of a daily 1-hour literacy lesson. However, teachers still spent an overwhelming amount of time on basic coding and decoding skills. Although there is no documented standard in Zambia for the amount of time in a 1-hour lesson that should be spent on these higher-order activities, it is clear that the current allotment is insufficient for the majority of learners to develop the ability to read fluently with comprehension.

Exhibit 46: Average Number of Intervals Criteria Were Performed – Comparison of Midline to Endline



* Denotes significance at 5 percent.

Domains: I-Orientation to Print; II-Phonemic Awareness and Phonics; III-Passage/Story Reading; IV-Reading Comprehension; V-Oral Language; VI-Writing

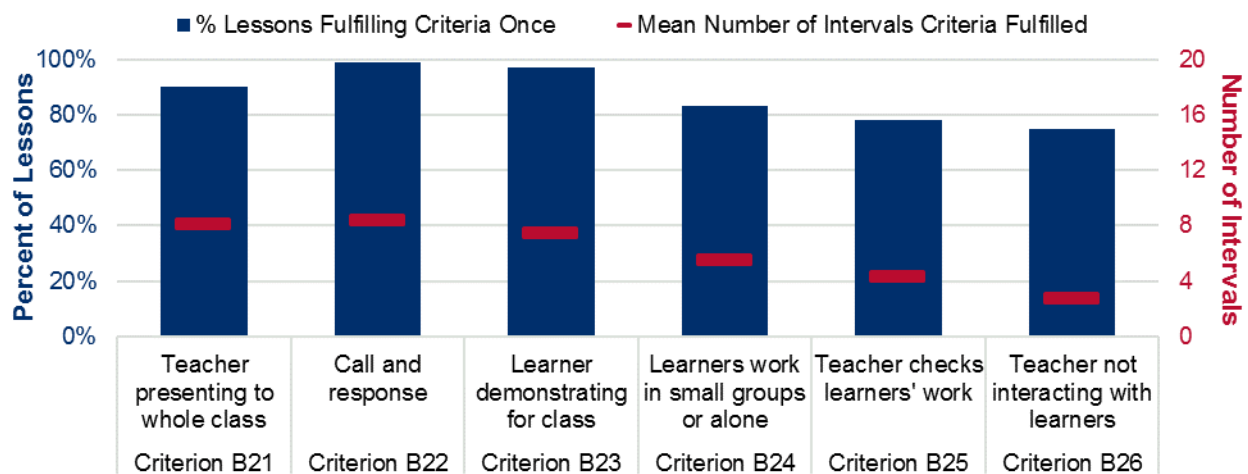
13

Teachers were observed using a variety of pedagogical approaches over the course of a literacy lesson, with a good amount of time spent with active learning methods.

Observed instructional practice has shifted since midline, and teachers were observed using a variety of pedagogical approaches during endline lessons. [Exhibit 47](#) illustrates general classroom practices in terms of whether the teacher was observed using this practice at least once during the observed 60 minutes, and the percentage of observed intervals in which the teacher used the practice (see [Exhibit 73](#) in Annex 2 for full results).¹⁰ Almost all teachers were observed using call and response, demonstrating to the whole class, and having learners demonstrate. Slightly fewer, but still at least three-quarters of observed teachers, were observed checking learners' work and having learners work in small groups. Teachers spent most classroom time instructing via call and response and teacher or learner demonstration, and less time having learners work in small groups or checking learners' work, but they used a variety of methods over the hour, with a good amount of time with these latter active learning methods. Three out of four teachers were observed not interacting with learners at some point during the lesson, but this comprised only 13 percent, on average, of observed intervals.

Exhibit 47: Observed General Classroom Practices at Endline

(n=105 classes)



During endline interviews, teachers reported on the frequency with which they conducted various activities during their literacy lessons. The most common practices that teachers reported doing five times a week were asking learners to copy text from the board (64 percent); asking learners to read aloud to the teacher or classmates (56 percent); having learners repeat after the teacher the sentences in a text (53 percent); modeling with a finger/pointer which direction to read (52 percent); and helping learners identify the sound each letter (or combinations of letters) in the alphabet produces (52 percent). Activities teachers most frequently stated that they “never” engaged in were inviting students to read texts/stories not in their textbooks (16 percent); asking learners to predict what will happen next in a story (13 percent); asking learners to read at home as homework (12 percent); and asking learners to identify whether there are any similarities between the events in a story and their own life experiences (12 percent).

¹⁰ Observation data on lesson delivery are not available for midline.

14

Head teachers' understanding of the need to use the MOGE-designated language of instruction in early primary grades has increased since midline.

According to Zambia's revised curriculum, the language of instruction for grades 1–4 in all learning areas is a familiar language, and English is the official language of instruction starting in grade 5. Head teachers were asked, according to the revised curriculum, in which grade should schools start to teach literacy in English. At endline, 50 percent of head teachers answered grade 5, compared to 36 percent at midline. In addition, there was a significant increase from midline to endline in the number of head teachers who reported that the language of instruction should be used all of the time in grade 1 (increasing from 78–92 percent, $p=0.033$) and grade 2 (increasing from 50–86 percent, $p<0.001$). This indicates a growing awareness of MOGE-supported approaches to literacy.

At endline, head teachers reported that the factors that most frequently prevented teachers from using the official language of instruction more often to teach literacy were not knowing how to teach reading in the language (64 percent), the school's not having local language materials (43 percent), and teachers thinking English is more important (41 percent). This appears to be a shift from midline, when head teachers' most common answer was that all grade 1 and 2 lessons were being taught in local language (29 percent).

15

A majority of teachers report using TTL-supported classroom management practices and MOGE-supplied resources.

Across the board, teachers interviewed reported using a broad variety of TTL-supported classroom management practices and predominantly MOGE-supplied textbooks and teachers' guides. The majority of teachers (83 percent) noted that the teaching approach they used the most often in their classroom was the Primary Literacy Program. Ninety percent of both male and female teachers reported having a scheduled time for teaching literacy, and 83 percent were able to produce a timetable that supported this. Eighty-seven percent of teachers claimed to teach literacy four to five times per week.

Eighty-two percent of teachers interviewed at endline said they develop daily literacy lesson plans (90 percent of these teachers could produce the lesson plan). The most frequently cited materials used to plan literacy lessons were MOGE learners' books (74 percent) and the teachers' guide (61 percent). In the classroom, teachers were observed using chalkboards (100 percent), government textbooks (69 percent), exercise books or slates (75 percent), and materials they made themselves (31 percent). Teachers confirmed that these were the main resources they had access to. A large majority of teachers (88 percent) said they had made their own teaching and learning materials, and 74 percent were able to produce them during the interview. Alternatively, a much smaller number of teachers had access to flashcards (46 percent), posters or "talking walls" (40 percent), supplementary reading materials (38 percent), and the TTL Stepping Stone phone (15 percent). This contrasted with 67 percent of head teachers reporting there was a Stepping Stone phone at the school, as well as some type of TTL material, such as supplementary readers, story cards, and flashcards.

Although the majority of teachers (93 percent) produced an attendance register when asked for it and 76 percent of teachers claimed to have taken attendance in the past week, more than half (64 percent) did not take attendance using the register during the observed class. Teachers' responses indicated a



Most teachers are incorporating TTL- and MOGE-promoted classroom practices and school management approaches related to literacy into their work.

broad variety of reasons for this, including, in some cases, the fact that data collection had disrupted the normal classroom routine.

Teachers stated that their most common course of action for repeated learner absences was to contact the learner's parents (76 percent), ask classmates about the learner (73 percent), or notify the head teacher (58 percent). Female teachers were significantly more likely to notify the head teacher (70 percent compared to 63 percent of male teachers) and contact the learner's parents (86 percent compared to 42 percent).

Nearly all teachers (99 percent) reported monitoring their learners' progress in learning to read, and 68 percent reported doing so daily or weekly. The most frequently cited methods of monitoring progress were asking questions about the lesson orally (66 percent), giving 5th/10th/13th-week assessments (65 percent) and listening to individual learners read aloud (62 percent). A significant number of female teachers tended to do these activities more frequently than their male colleagues. Both male and female teachers rarely used the red-level tracker (20 percent) and continuous assessment (32 percent), and less often observed learners in group activities (43 percent) and asked learners to tell about what they have just read (49 percent). Different teachers appeared to conduct recordkeeping related to learners' progress at different frequencies (daily, weekly, monthly, etc.), and records available at the time of interview reflected that variation.

16 | **Head teachers and observed teachers differed in their responses regarding the number of times literary classes were monitored by head teachers.**

There was no significant difference from 2014–2016 in the proportion of head teachers who reported monitoring teachers (from 79 percent at midline to 83 percent at endline). Head teachers paid by the government are significantly more likely to observe their teachers in the classroom—93 percent, compared to 74 percent of other types of head teachers, at both endline and midline. On average, head teachers reported monitoring classes 1.8 times at midline and 2 times at endline.

Teachers at midline and endline reported similar amounts of monitoring by head teachers. Endline teachers reported receiving an average of four visits and midline teachers an average of five visits. Approximately 23 percent of teachers at both midline and endline reported not receiving any monitoring by their head teacher.

At endline, head teachers reported holding regular meetings on literacy. Thirty-four percent reported meeting monthly, 26 percent reported meeting every 2 weeks, 23 percent reported meeting once per term, and 3 percent reported meeting once per year. The remaining respondents either reported never meeting or did not respond.

17 | **Teachers show attitudes conducive to good practices in literacy instruction and affirm all children's ability to learn.**

At midline and endline, teachers were asked to respond (agree, disagree, or no opinion) to a number of statements assessing teacher attitudes and beliefs toward their learners and the teaching of literacy. For the most part, teacher attitudes did not change significantly between midline and endline, and there was no significant difference between male and female teachers at endline. The vast majority of teachers reaffirmed their belief that all learners can learn to read (94 percent) and write (97 percent).

A few teachers believed that learners have a lot of difficulty learning to write (53 percent) and that it is difficult for learners to learn to read (36 percent). Teachers were ambivalent in their opinions of whether girls learn to read or write faster than boys. However, when asked whether boys learn to read or write faster than girls, the majority of teachers disagreed (71 percent). Approximately 12 percent of respondents

had no opinion on the matter. Approximately three-quarters of teachers continued to assert that girls like to read (73 percent). Fewer believed that boys like to write (58 percent).

Thirty-six percent of teachers asserted that it is better to teach reading and writing as separate subjects, while 77 percent of teachers agreed that it is better to teach reading and writing together. However, there was a contradictory response to the statement that learners should learn to read before learning to write (61 percent). Close to half of teachers asserted that learners must memorize a text before they can understand it (53 percent). There was near universal agreement around the statement that telling stories helps create interest in reading (95 percent).

Taken on the whole, these attitudes generally reflect equitable beliefs in different children's ability to read and attitudes that are conducive to the teaching of literacy according to the MOGE curriculum and the Primary Literacy Program.

5.4. FACTORS ASSOCIATED WITH LEARNER OUTCOMES

18 Practicing reading in school and at home, lower learner-teacher ratios, and stronger oral language comprehension are key factors associated with higher reading outcomes, which affirms the ability of targeted interventions to further improve early grade reading achievement.

Previous sections of the report look at learner performance, MOGE support, and teacher practice and find all three to have made significant and practically important improvements over the course of TTL. Section 4 provides a range of school-level summary descriptive statistics that are often found to be important factors influencing educational quality. Education research frequently finds a range of learner-level characteristics, such as socioeconomic status, sex, and rural or urban settings, to affect learning outcomes, but these characteristics are typically ones over which actors have little control. Thus, it is useful to know the extent to which vulnerable groups are at risk, identify conditions under which negative effects can be lessened, and pinpoint factors associated with strong learning outcomes.

To test these questions, three linear regression models were fitted for three separate outcome variables representing three subtasks on EGRA: letter sound knowledge (task 2), non-word reading (task 3), and oral passage reading (task 5a). These models include phase (midline and endline) as a key variable, and the analysis process tested a variety of characteristics of learners and their households, school characteristics, and teacher practice. Most learner background variables were initially tested for their association with the three selected outcome variables, but only variables that showed a statistically significant relationship with the outcome variables were retained in the final models. [Exhibit 48](#) presents the final linear regression models, including coefficients and standard errors for the three EGRA tasks.

Exhibit 48: Variables Associated with Learner Performance

Variable	Letter Sounds		Non-word Reading		Oral Passage Reading	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Evaluation Phase						
Midline (r)	2.550**	0.834	1.173**	0.453	1.683**	0.536
Endline						
Practices reading with somebody at home						
No (r)	2.061***	0.562	1.638***	0.495	2.051**	0.729
Yes						

Variable	Letter Sounds		Non-word Reading		Oral Passage Reading	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Read on own outside of school						
Never						
Less than monthly						
Monthly	0.981***	0.279	0.920***	0.154	1.066***	0.173
Weekly						
Daily						
Number of days practicing reading in school per week Range: 0–5 days per week	0.755***	0.216	0.340**	0.125	0.407**	0.151
School Setting						
Rural (r)	–4.641***	1.342	–2.530***	0.742	–3.219***	0.888
Urban						
Learner-to-teacher ratio Number of learners per teacher	–0.033*	0.015	–0.020*	0.008	–0.020*	0.009
Task 1 (listening comprehension): answers correct Range: 0–5	1.647***	0.208	1.106***	0.177	1.466***	0.246
Constant	13.244***	3.250	7.962***	1.840	8.143***	2.189
R-squared	0.227		0.217		0.210	

Note: Statistical significance denoted by * $p \leq 0.05$, ** $p \leq 0.01$, and *** $p \leq 0.001$
(r) denotes reference category for dichotomous variables.

In [Exhibit 48](#), coefficients indicate the amount of change in the outcome variables associated with a one-unit increase in the variable measuring the coefficient; changes can be positive or negative. Outcome variables are measured in letters, non-words, and words per minute, respectively, for these three tasks. Thus, a one-unit increase in the coefficient is associated with an increase or decrease by the amount of the coefficient in the number of letters or words learners read in 1 minute. For example, practicing reading in school 1 day per week, as compared to 0 days per week, is associated with a child reading 0.755 additional letters correctly in a minute, an additional 0.340 non-words, and an additional 0.407 words. Some variables in the model are categorical, such as school setting (which is either urban or rural). In this case, the coefficient indicates that being associated with the second group results in the increase or decrease specified by the coefficient. Continuing with school setting as an example, children in urban schools read 4.641 *fewer* letters correctly per minute, 2.530 fewer non-words, and 3.219 fewer words; the fact that the result is *fewer* letters or words per minute can be seen in the negative sign preceding the coefficient.

These preliminary regression results indicate that learner performance has improved since midline across all three EGRA tasks analyzed, even when controlling for the variables listed in [Exhibit 48](#). Three different variables that measure practicing reading were all predictive: practicing reading with someone at home, reading on one's own outside of school, and the number of days during which reading is practiced at school. Being in an urban school as well as having a higher learner-to-teacher ratio; both had negative effects on learner performance, indicating that rural learners and those at less crowded schools perform better. The model includes oral language comprehension (EGRA task 1), and this is an important contribution to the model, increasing its predictive value of the outcome (measured by r-squared) from about 16–21 percent or greater. All the independent variables are associated with changes in outcomes in the logical direction (e.g., children who practice reading more also perform better on the reading skills measured by EGRA, as would be expected).

Notably, many innate learner characteristics that schools cannot alter do not surface as significant predictors of learner performance. Specifically, a child's sex, age, socioeconomic status, and whether the language of instruction is spoken at home are not significant predictors of reading ability in the final models. Age dropped out as a significant variable when school location was added to the model (rural community learners are about 0.4 years older, on average, than urban community school learners), though when school setting was not included in the model, age was a positive predictor of performance. Children who reported speaking a language other than the language of instruction at school or home did not perform significantly worse in these models. There were also a number of school-level factors that were not predictive of reading outcomes, including school size, number of forms of support to the school (as a proxy for school resources), and head teacher employment status. The frequency of PCSC meetings and employment status of the grade 2 teacher was predictive in most models, but interestingly became insignificant when listening comprehension was included in the model. Classroom practice, as measured by the classroom observation protocol, did not correlate with learning outcomes, and thus was not tested in these models. However, this is probably a limitation of the tool, the primary purpose of which was to measure what teachers *are capable of* and not necessarily what their *routine practices* are, which would require multiple assessments throughout the year.

In many cases, these results align with regression models resulting from the 2014 Grade 2 National Assessment, which bolsters the strength of this evidence. Factors where there is alignment across these two studies may also indicate areas where the effects described above are not specific only to community school learners.

While these models do not imply a causal relationship, overall, these results do indicate a number of areas that are amendable to intervention and are positively associated with reading outcomes. Moreover, a number of innate learner characteristics that are beyond schools' control do not appear to be significant predictors of reading ability in community school settings, though these data cannot compare how these learners perform in comparison to learners in other school types in Zambia. Finally, by controlling for other potential explanatory factors and repeatedly finding the phase of the evaluation to be a significant and positive predictor of reading ability, these models offer further evidence that reading scores have truly improved over the course of TTL and that these improvements are not merely due to changes in underlying factors.

6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The first part of this section summarizes the answer to each of the three evaluation questions (see [Exhibit 5](#)). This summary is followed by a set of conclusions that seek to draw deeper meaning from the combination of results and in light of other evaluation data collected over the life of TTL. The section closes with a set of recommendations related to strengthening early grade literacy performance of community schools and their learners.

6.1. SUMMARY RESPONSES TO EVALUATION QUESTIONS

EVALUATION QUESTION 1: To what extent have TTL interventions improved early grade literacy achievement among male and female learners in community schools across six provinces, compared to baseline?

The overall reading ability of grade 2 learners in community schools has increased significantly since baseline. Although very few grade 2 learners are reading at grade level in 2016, many of the improvements in specific EGRA tasks are large in practical terms, as measured by effect sizes of the change from baseline to endline. This improvement is particularly apparent for the fundamental reading competencies of phonics, decoding, and reading fluency, with smaller improvements for reading comprehension. Performance shows no consistent statistical differences between male and female learners across the three evaluation phases, meaning that literacy skill acquisition has not come at the expense of either sex.

In almost all instances, the direction of progress has been consistent, with midline mean scores lying between baseline and endline means and the rate of improvement from midline to endline appearing larger compared to change from baseline to midline. This increase is logical, given the timing of TTL intervention activities and children's progression through school: learners assessed at midline (late 2014) had already begun grade 1 before their teachers received the first TTL training in mid-2013 and could begin to implement any changes in teaching practice. By comparison, learners assessed at endline commenced grade 1 in January 2015, by which time their teachers had had the opportunity to participate in several TTL training courses and incorporate new practices into their classrooms.

Listening comprehension skills have remained unchanged since midline (baseline data are not comparable), indicating that oral language proficiency may be more difficult to change. Weak oral language ability may be impeding development of other literacy skills. The combination of weak reading comprehension scores and weak listening comprehension scores suggests that learners may be lacking comprehension strategies in general, both in reading and listening. Zambia community school catchment areas often contain learners whose language of instruction at school is not the language they use at home, in church, at the market, and even at play (MOGE 2013). Bivariate analysis indicates that learners who spoke the language of instruction at home performed moderately better on listening comprehension in the language in which they were tested, although this association explains only a small portion of the weak performance in listening comprehension. The Zambia EGRA instruments do not capture underlying linguistic and cognitive skills that might further explain listening comprehension results.

Progress is not consistent across all provinces. Central and Copperbelt have shown improvement in performance across all EGRA tasks at endline, whereas other provinces have more mixed results across tasks. While there is considerable variation both across and within schools at midline and at endline, improvement in literacy skills from midline to endline is still evident. At midline, about 45 percent of sampled learners attended a school whose task 2 median score was zero; at endline, about one-third of learners did, and another third attended a school whose task 2 median was greater than the overall mean of eight letters per minute.

EVALUATION QUESTION 2: How has the MOGE's engagement with community schools changed since baseline?

The MOGE has significantly increased the proportion of community schools it is supporting since baseline, in terms of teaching and learning materials, free basic materials, teacher training, monitoring, deployment of government teachers, and infrastructure. The only area in which the proportion of schools receiving MOGE support did not increase since the baseline is direct financial support.

Community schools are also receiving more *types* of support. The quality of monitoring support that MOGE staff provides appears to be improving, with more visits, including classroom observation with feedback. The 2014 OGCS clarified that allocation of government grants to community schools to contain (1) a minimum amount of support to be allocated to both government and community schools and (2) another portion to be distributed based on cost drivers, such as enrollment, number of teachers, distance from District Education Board Secretaries' offices, and terrain (MOGE 2016b: 36). More remote sampled schools have received more grant support, indicating that this policy is being implemented. The presence of a government-seconded head teacher is also associated with more types of MOGE support for a community school. Despite these improvements in MOGE support, however, community school head teachers still report inadequate resources at their schools. The improvement in perceptions of support between head teachers and District Education Board Secretaries since baseline further indicates increased MOGE responsibility toward community schools and their further integration into the Zambian education system. At endline, the MOGE was reported as the primary provider of support among sampled schools.

EVALUATION QUESTION 3: To what extent are male and female teachers implementing TTL-supported literacy teaching methods?

Observed classroom practices are shifting to strengthened literacy teaching methods, with lessons becoming more balanced across the five literacy competencies. Compared to midline (no comparable data at baseline), there are significant observable changes in classroom practice, with more time spent on higher-order literacy skills, and teachers reporting these same changes in their classroom practices. Rather than focusing exclusively on letter sounds or basic decoding skills, more teachers were observed incorporating reading comprehension and fluency activities. At the same time, although the amount of lesson time focused on these higher-order skills has increased, there remains an imbalance between activities focused on basic phonics and those focused on developing oral and written comprehension skills.

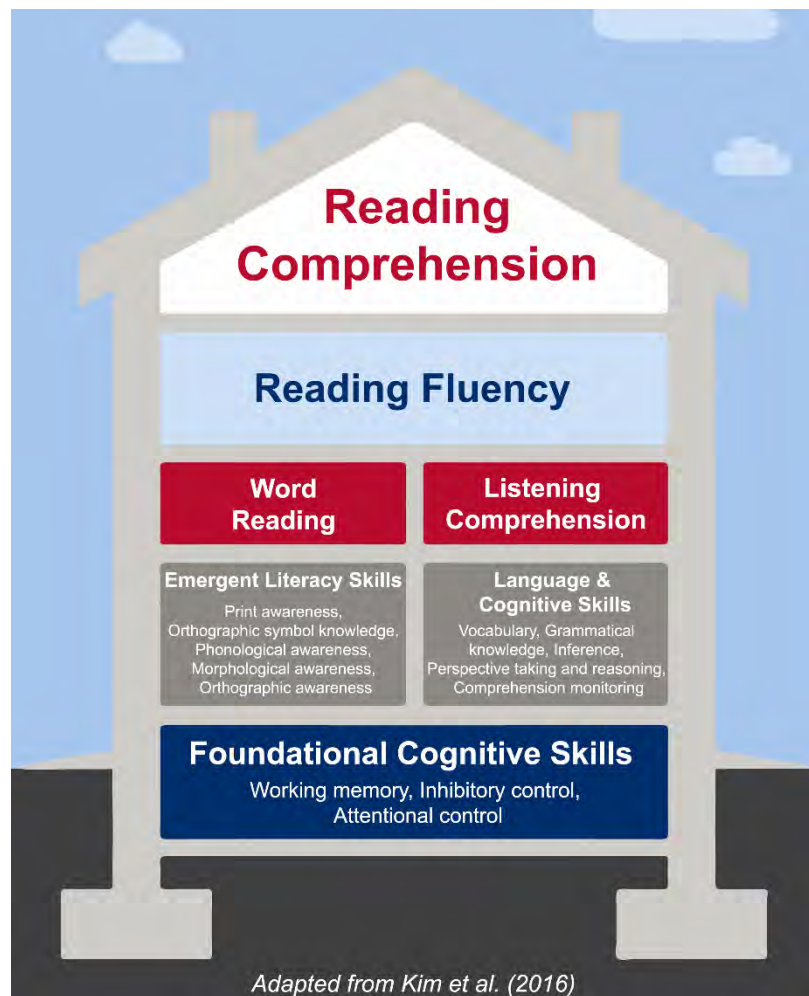
Teachers were observed using a variety of pedagogical approaches over the course of a literacy lesson, with a good amount of time spent with active learning methods. Most teachers reported using TTL-supported practices and MOGE materials and showed greater understanding of the MOGE-designated language of instruction. Reported practices indicate greater use of TTL- and MOGE-promoted pedagogical approaches and classroom and school management practices.

6.2. CONCLUSIONS

Early grade reading acquisition is a complex process requiring multiple linguistic and cognitive skills (Kim et al. 2016). As [Exhibit 49](#) illustrates, reading comprehension—the USAID target and the ultimate end goal of learning to read—requires the learner to be able to decode or recognize words (word reading) and comprehend oral language at the discourse level. These two abilities are key component skills that, when occurring fluently (accurately and with speed), enable reading comprehension. Each of these skill areas is built on a complex set of foundational skills: emergent literacy skills, language and cognitive skills, and foundational cognitive skills.

In examining the endline results and drawing deeper meaning, it is clear not only that there is complexity in literacy acquisition, regardless of the context, but also that Zambia, and particularly its community schools, present their own challenges to achieving grade-level reading performance.

Exhibit 49: Component Skills in Reading Comprehension and Their Structural Relationships



The context of improving learner performance is challenging in Sub-Saharan Africa, in Zambia, and particularly in community schools, influencing the possible speed of change.

In the years before the MOGE's policy shift in literacy instruction, reading scores in Zambia and much of Southern Africa had been declining. The Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) testing consortium showed grade 6 reading levels declining in Zambia from 1995–2007, the most recent year for which SACMEQ data are available (Musonda and Kaba n.d.). Five other countries in the region also showed reading score declines over the same period. A 2013 review of SACMEQ data found that the median rate of growth in the region was 0.0115 standard deviations per year, which is the equivalent of an annual effect size of 0.0115 (Beatty and Pritchett 2012). These trends show that some of the challenges to literacy acquisition are national and even regional in scope.

The context of community schools poses further challenges for learner performance. *Ensuring Inclusive and Quality Education for All: A Comprehensive Review of Community Schools in Zambia* (Frischkorn and Falconer-Stout 2016) found that most community schools remain severely under-resourced in terms of basic educational inputs, even though many also benefit from community commitment and volunteer teachers' dedication. The results from the endline evaluation indicate that the average teacher-to-learner ratio is 1:60 and that more than half of community school teachers have no professional qualifications. Although resource availability is slowly increasing, with the MOGE providing more resources to more community schools, it is not yet addressing the full complement of needs.



In this context, TTL, with USAID and the MOGE, worked to improve reading, and to reverse a previous negative trend. The extent of improvement observed in the TTL intervention area over the past 4 years, while not meeting targets, is not only strong, but also historically remarkable, reaching as high as 14 times the annual regional median rate of growth observed by Beatty and Pritchett (2013) for the oral passage reading task in Central Province.

2 The Zambian policy context for improved literacy practice has begun to extend to community schools (with TTL assistance), and the MOGE's commitment and capacity to support community schools has increased. Targeted and consistent inputs are needed to ensure the delivery of high-quality early grade education at all community schools.

Under the Primary Literacy Program, the MOGE has instituted a broad array of changes that affect literacy teaching and learning in early grades across all school types. Many of these policy changes align with USAID-supported, evidence-based approaches to improving literacy.

TTL has been an important actor in the development and implementation of these policy changes, ensuring the inclusion of the community school perspective. The Zambian government's acknowledgement and incorporation of community schools into the education system has increased significantly since baseline, as seen by the sheer number of MOGE resources received by each school, formal policy declarations such as the OGCS, the way MOGE conducts community school monitoring, and changes in MOGE-approved teaching practices in classroom. This is further demonstrated in teachers' reported reliance on MOGE-supplied resources, such as learners' books and teacher's guides. However, these inputs have yet to reach the level of material need to ensure educational quality across the diverse landscape of Zambia's community schools.

3 Teachers' practices are changing to align with TTL- and MOGE-supported standards, mirroring the changes learner performance: Stronger learner results have been observed in areas teachers report (and are observed) practicing the most during lessons (phonics and decoding), and there is less progress in areas teachers practice less (oral language development and reading comprehension).

The changes evident in teaching practices at endline reflect a continuation of a trend observed during the performance evaluations in project year 2 (2013) and year 4 (2015) and at midline (2014). In year 2, teachers were first observed beginning to switch to a phonics-based approach to literacy, a trend that was confirmed by the following year's midline evaluation. An imbalance was noted between efforts devoted to lower-order versus higher-order literacy skills, but the phonics-based approach was still new. In year 4, teachers noted several factors that had helped them make changes, but also described many challenges they face in incorporating teaching these higher-order skills in the classroom.

Performance evaluation data from 2015 indicated that their proficiency in the language of instruction is a challenge to many teachers' ability to teach literacy (Frischkorn, Falconer-Stout and Messner 2016). This again mirrors broader regional trends; in Kenya, for example, the mismatch between teacher deployment strategies and language instruction policy has been cited as a factor inhibiting mother-tongue instruction (Piper, Zuilkowski, and Ong'ele 2016: 778).

In both the year 2 and year 4 performance evaluations, teachers stated that they were changing specific classroom practice as a result of TTL training, and these same changes were observed during classroom observation in the midline and endline evaluations. The emergence of this theme at each evaluation phase, confirmed by qualitative and quantitative data sources, provides strong evidence supporting the positive role TTL has played.

The areas teachers were observed to focus the most during lessons—competencies related to phonics and decoding—are the areas in which EGRA task performance has been the strongest. At endline, there was a significant, though still limited, increase in focus on higher-order literacy skills such as reading connected text and comprehension, and grade 2 teachers benefiting from TTL training and MOGE support were incorporating more appropriate pedagogical approaches and focusing somewhat more on oral language development and reading comprehension.



4 Learner performance is improving, but not quickly enough to meet MOGE targets. More research is needed to understand how to leverage factors influencing performance at both learner and school levels.

Learning performance shows improvements at endline, both over baseline and midline. Based on the current rate of change (2012–2016), community schools are not improving reading outcomes at a pace sufficient to meet the MOGE's 5-year targets (to be achieved by 2019) for the reader and emergent reader categories. Yet, comparing the TTL midline with the 2014 Grade 2 National Assessment EGRA data also reveals that community schools started at a marginally lower point than national estimates, which combine all four Zambian school types (RTI International 2015). In addition, the substantial variation in reading outcomes within and between schools reveals that future community school policy and support cannot take a one-size-fits-all approach to increase performance at all schools along the spectrum and meet performance targets.

Examining the framework in [Exhibit 49](#), the ongoing weakness in listening comprehension in the language of instruction could be hindering progress in acquisition of reading fluency and reading comprehension. Zambia is not alone in this respect; globally, the evidence base is weak regarding effective ways to facilitate listening comprehension (Kim et al. 2016).

Regression analysis confirms this link between listening comprehension and other reading tasks, as well as the importance of learners practicing reading alone and with others, both in the classroom and at home. Sex and socioeconomic status showed no significant predictive value. Greater understanding of the factors driving higher performance at rural schools is needed; factors that may play a role include language diversity in urban areas, age (maturity of students), and school size.

In summary, TTL interventions, as laid out in its development hypothesis ([Exhibit 8](#)), have shown that it is possible to improve support to community schools and classroom practice and to ultimately improve reading scores among community school learners. To help maintain and expand emerging progress, the next section explores considerations for policymakers, donors, and implementers working with Zambia's community schools.

6.3. RECOMMENDATIONS

1 MOGE and its partners should target community schools in ways that recognize the wide variation within and among the schools.

Although community school reading outcomes are improving on the whole, viewing them overall, or even at the provincial level, masks the substantial variation that exists both within and between these schools. More work will be required to determine how to simultaneously increase performance at all community schools. Nevertheless, it should be expected that schools at different points along the performance spectrum require very different policy responses, meaning that community school policy and support cannot take a one-size-fits-all approach. Top-performing schools require motivation and encouragement to continue, and even strengthen, their good work, and they should be studied further to identify best practices that can be promoted at other schools. Schools where all learners are performing below average need more intensive support and mentoring. Schools with a wide diversity in learner performance need support to identify factors behind the uneven performance and provide remedial instruction to learners who are falling behind.

2 USAID and others should support additional research to explore factors associated with weak oral language proficiency and ways it may be impeding the development of other literacy skills among community school learners.

Weak listening comprehensions scores in the language of instruction indicate that learners may not fully comprehend the content of literacy lessons, slowing or limiting the absorption of necessary literacy skills. More directly, weak listening comprehension indicates broader challenges of comprehension and means that increasing oral passage reading fluency is likely not enough on its own to improve reading comprehension. Improving listening comprehension is an area where the global evidence is weak, and Zambia's multilingual environment is a challenging context in which to address this weakness. Additional research in Zambia is necessary to understand learners' specific linguistic contexts and the impact these contexts have on the learning environment and on the means to overcome unfamiliar languages of instruction in the classroom.

- 3** The MOGE should continue its integration of community schools into the Zambian education system and provide consistent distribution of resources, in line with the level of material need, to ensure educational quality across the diverse landscape of community schools.

Community schools have consistently delivered education to male and female learners from lower socioeconomic groups, and they play an important role in Zambia's educational landscape. The MOGE has acknowledged this role with increasing proportions of support and fuller integration at district and zonal levels. Although resource availability is increasing to community schools, it has yet to address the full complement of needs, including dilapidated or nonexistent infrastructure, limited teaching and learning materials, and untrained volunteer teachers. District Education Boards can ensure continued and consistent support to community schools through equitable distribution of textbooks, basic classroom materials, and building materials for infrastructure development; inclusion of community schools' head teachers and teachers in CPD activities; and regular monitoring of school and classroom activity to support high-quality education for all of Zambia's children. At the same time, as other studies have noted (Frischkorn and Falconer-Stout 2016; Falconer-Stout, Eunifidah and Mayapi 2014; Falconer-Stout and Kalimaposo 2014), care should be taken that this continued integration does not minimize the unique strengths of community schools, particularly the community oversight and the dedication of volunteer teachers.

- 4** The MOGE and its partners should provide teachers with additional guidance on structuring literacy lessons and integrating higher- and lower-order literacy skills into teaching practice in a way that produces the desired reading levels in their learners.

Although literacy lessons are becoming more balanced across the five literacy competencies and are including more of the higher-order literacy skills, community school teachers need more support and guidance to continue making changes. There is currently no documented standard in Zambia on how teachers should allocate lesson time to higher-level literacy activities, but current instructional practices observed in this evaluation are clearly not yet sufficient for the majority of learners to develop the ability to read fluently with comprehension. Effort should be directed at providing teachers with additional guidance on the design and implementation of lessons, as well as practical ways to effectively negotiate the particular challenges found in community schools, such as high learner-to-teacher ratios and multilevel classrooms. Several of the specific teaching challenges were documented in TTL's year 4 performance evaluation (Frischkorn, Falconer-Stout and Messner 2016), along with techniques that teachers have found to be successful. To ensure the sustainability of future improvements, the turnover rates of volunteer teachers in community schools should also be considered in continuing to strengthen durable systems for ongoing CPD.

ANNEX I. DETAILED EGRA RESULTS

This annex presents complete EGRA results for all seven tasks across the three evaluation phases. The results are presented for three different levels of aggregation: province, language, and sex. Each task is presented in a separate exhibit for each level of aggregation; the full exhibit list is provided below.

Each exhibit presents two sets of statistics: 1) mean score with standard errors and 95 percent confidence intervals, as well as the p-value and effect size for the comparison of the baseline and endline means; 2) the distribution of learners' scores across the four MOGE performance categories ("benchmarks"), as well as the p-value for the comparison of the baseline and endline distributions. The statistical test for comparison of means is an independent t-test, and for comparison of distributions is a chi-square test of homogeneity.

The benchmarks for tasks 3, 5a, and 5b, are taken from the MOGE benchmarks for early grade reading (MOGE 2015a). The categories for tasks 1 and 6 are aligned to task 5b, because the tasks are scored in a similar manner and on the same scale. Task 4 has only three questions; a separate category is presented for each possible total score.

EGRA RESULTS BY PROVINCE

[Exhibit 50: Task 1 – Listening Comprehension:
Zambian Language of Instruction](#)
[Exhibit 51: Task 2 – Letter Sound Knowledge](#)
[Exhibit 52: Task 3 – Non-word Reading](#)
[Exhibit 53: Task 4 – Orientation to Print](#)
[Exhibit 54: Task 5a – Oral Passage Reading](#)
[Exhibit 55: Task 5b – Reading Comprehension](#)
[Exhibit 56: Task 6 – Listening Comprehension:
English](#)
[Exhibit 57: Task 7 – English Vocabulary](#)

EGRA RESULTS BY LANGUAGE

[Exhibit 58: Task 1 – Listening Comprehension:
Zambian Language of Instruction](#)
[Exhibit 59: Task 2 – Letter Sound Knowledge](#)
[Exhibit 60: Task 3 – Non-word Reading](#)
[Exhibit 61: Task 4 – Orientation to Print](#)
[Exhibit 62: Task 5a – Oral Passage Reading](#)
[Exhibit 63: Task 5b – Reading Comprehension](#)
[Exhibit 64: Task 6 – Listening Comprehension:
English](#)
[Exhibit 65: Task 7 – English Vocabulary](#)

EGRA RESULTS BY SEX

[Exhibit 66: Task 1 – Listening Comprehension:
Zambian Language of Instruction](#)
[Exhibit 67: Task 2 – Letter Sound Knowledge](#)
[Exhibit 68: Task 3 – Non-word Reading](#)
[Exhibit 69: Task 4 – Orientation to Print](#)
[Exhibit 70: Task 5a – Oral Passage Reading](#)
[Exhibit 71: Task 5b – Reading Comprehension](#)
[Exhibit 72: Task 6 – Listening Comprehension:
English](#)

EGRA Results by Province

Exhibit 50: Task 1 – Listening Comprehension: Zambian Language of Instruction¹¹

Task 1: Listening Comprehension: Zambian Language of Instruction (3 Questions at Baseline; 5 Questions at Midline and Endline)																		
Province	Mean percent questions correct Standard Error of Mean (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 questions correct)			% pre-emergent (1 question correct)			% emergent readers (2–3 questions correct)			% readers (4–5 questions correct)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
Central	50% 0.199 (37%–63%)	50% 0.152 (44%–56%)	52% 0.193 (44%–60%)	0.523	0.03	25%	12%	10%	28%	13%	19%	26%	50%	38%	20%	26%	33%	0.009
Copperbelt	67% 0.254 (50%–87%)	42% 0.156 (36%–50%)	52% 0.246 (42%–62%)	0.130	0.21	5%	14%	11%	35%	27%	16%	12%	40%	48%	48%	19%	25%	0.003
Eastern	57% 0.172 (47%–70%)	62% 0.134 (56%–66%)	68% 0.078 (66%–72%)	0.023	0.15	11%	5%	4%	35%	8%	3%	23%	47%	34%	32%	40%	59%	<0.001
Lusaka	53% 0.197 (40%–67%)	64% 0.083 (60%–68%)	68% 0.119 (64%–72%)	0.171	0.13	28%	3%	2%	24%	4%	3%	20%	51%	49%	27%	42%	46%	0.673
Muchinga	60% 0.263 (43%–77%)	42% 0.176 (34%–50%)	38% 0.149 (32%–44%)	0.314	–0.08	31%	16%	19%	7%	26%	23%	15%	37%	44%	48%	20%	14%	0.091
Southern	90% 0.104 (83%–100%)	42% 0.135 (36%–46%)	46% 0.157 (40%–52%)	0.337	0.07	1%	31%	14%	2%	12%	17%	19%	31%	39%	77%	26%	29%	<0.001

¹¹ Baseline scores for Task 1 are not comparable to midline and endline scores (see Annex 3); these scores are presented for informational purposes only. Reported p-values and effect sizes are for comparison of midline and endline scores.

Exhibit 51: Task 2 – Letter Sound Knowledge

Task 2: Letter sound knowledge (100 letters)																		
Province	Mean letters per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 letters per minute)			% pre-emergent readers (1–15 letters per minute)			% emergent readers (16–40 letters per minute)			% readers (41+ letters per minute)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
Central	2.2 <i>0.564</i> (1.1–3.4)	9.3 <i>0.998</i> (7.2–11.3)	11.1 <i>1.067</i> (8.9–13.2)	<0.001	1.13	71%	27%	24%	23%	56%	50%	5%	17%	26%	0%	0%	0%	<0.001
Copperbelt	2.2 <i>0.903</i> (.3–4.0)	5.2 <i>0.582</i> (4.0–6.2)	10.5 <i>2.370</i> (5.7–15.3)	0.003	1.12	73%	40%	31%	24%	56%	48%	3%	5%	20%	0%	0%	1%	<0.001
Eastern	3.5 <i>1.168</i> (1.1–5.9)	5.0 <i>0.701</i> (4.1–6.9)	9.2 <i>1.528</i> (6.1–12.2)	0.005	0.81	60%	38%	32%	31%	52%	42%	9%	11%	25%	0%	0%	0%	<0.001
Lusaka	4.2 <i>0.555</i> (3.1–5.3)	4.0 <i>1.456</i> (1.0–7.0)	5.2 <i>1.591</i> (2.0–8.4)	0.563	0.14	34%	63%	66%	60%	29%	24%	6%	8%	8%	0%	0%	2%	<0.001
Muchinga	4.1 <i>1.366</i> (1.3–6.9)	6.5 <i>1.231</i> (4.0–9.0)	8.3 <i>1.346</i> (5.6–11.1)	0.033	0.36	55%	32%	42%	39%	57%	40%	6%	11%	18%	1%	0%	0%	<0.001
Southern	0.5 <i>0.182</i> (0.1–0.8)	1.4 <i>0.257</i> (0.8–1.9)	6.4 <i>2.897</i> (0.5–12.3)	0.048	0.87	91%	77%	45%	8%	22%	45%	1%	1%	8%	0%	0%	2%	<0.001

Exhibit 52: Task 3 – Non-word Reading

Task 3: Non-word Reading (50 invented words)																		
Province	Mean words per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 words per minute)			% pre-emergent readers (1–14 words per minute)			% emergent readers (15–29 words per minute)			% readers (30+ words per minute)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
Central	0.46 0.297 (–0.1–1.1)	3.0 0.667 (1.6–4.3)	5.2 1.414 (2.3–8.1)	0.002	0.71	95%	71%	63%	3%	21%	18%	1%	7%	17%	0%	0%	2%	<0.001
Copperbelt	1.0 0.558 (–0.1–2.1)	2.2 0.640 (0.9–3.5)	4.8 1.462 (1.8–7.8)	0.021	0.72	90%	85%	70%	8%	11%	19%	2%	3%	10%	0%	1%	1%	<0.001
Eastern	1.6 0.788 (0.0–3.2)	2.1 0.503 (1.1–3.1)	4.4 1.268 (1.8–7.0)	0.070	0.54	84%	76%	70%	14%	18%	18%	2%	6%	10%	0%	0%	2%	<0.001
Lusaka	0.7 0.304 (0.1–1.4)	1.1 0.477 (0.2–2.1)	2.1 0.642 (0.8–3.5)	0.048	0.39	87%	89%	84%	12%	9%	13%	2%	2%	3%	0%	0%	0%	0.401
Muchinga	2.0 0.764 (0.4–3.5)	2.3 0.684 (0.9–3.7)	3.5 0.868 (1.8–5.3)	0.180	0.17	84%	76%	72%	11%	17%	17%	2%	7%	10%	2%	0%	1%	0.063
Southern	0.3 0.215 (–0.1–0.7)	0.6 0.244 (0.1–1.1)	2.4 1.494 (–0.7–5.4)	0.183	0.66	96%	92%	84%	3%	7%	11%	1%	1%	6%	0%	0%	0%	<0.001

Exhibit 53: Task 4 – Orientation to Print¹²

Task 4: Orientation to Print (3 questions)													
Province	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		p-value	Effect size	% scoring zero correct (0 questions correct)		% pre-emergent readers (1 question correct)		% emergent readers (2 question correct)		% readers (3 question correct)		p-value
	Midline	Endline			Mid	End	Mid	End	Mid	End	Mid	End	
Central	93% 0.050 (90%–97%)	90% 0.026 (87%–93%)	0.226	–0.08	3%	4%	3%	3%	5%	7%	89%	86%	0.607
Copperbelt	60% 0.117 (53%–67%)	90% 0.035 (87%–93%)	<0.001	0.44	31%	4%	13%	5%	11%	5%	45%	87%	<0.001
Eastern	77% 0.087 (70%–83%)	93% 0.078 (87%–97%)	0.004	0.30	18%	8%	3%	3%	12%	2%	67%	87%	<0.001
Lusaka	93% 0.056 (87%–97%)	93% 0.055 (90%–97%)	0.555	0.00	5%	5%	1%	3%	2%	6%	92%	86%	0.023
Muchinga	83% 0.083 (80%–90%)	83% 0.042 (87%–87%)	0.699	0.00	9%	9%	4%	6%	13%	13%	73%	72%	0.784
Southern	83% 0.891 (77%–87%)	93% 0.060 (90%–97%)	0.005	0.20	12%	2%	4%	4%	12%	12%	71%	82%	<0.001

¹² Task 4 was introduced during the final period of data collection at baseline; consequently, data are not available for Eastern and Lusaka provinces and sample sizes for the other provinces are too small to enable statistical comparison. Reported p-values and effect sizes are for comparison of midline and endline scores.

Exhibit 54: Task 5a – Oral Passage Reading

Task 5a: Oral Passage Reading (40–68 words, depending on EGRA language version and evaluation phase)																		
Province	Mean words per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 words per minute)			% pre-emergent readers (1–19 words per minute)			% emergent readers (20–44 words per minute)			% readers (45+ words per minute)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
Central	0.5 <i>0.328</i> (–0.1–1.2)	2.5 <i>0.546</i> (1.4–3.6)	5.7 <i>1.784</i> (2.05–9.3)	0.007	0.66	96%	71%	67%	3%	26%	19%	1%	2%	13%	0%	0%	1%	<0.001
Copperbelt	1.3 <i>0.849</i> (–0.4–3.1)	2.1 <i>0.596</i> (0.9–3.3)	5.0 <i>1.594</i> (1.8–8.3)	0.050	0.60	91%	83%	72%	5%	13%	22%	4%	3%	6%	0%	0%	0%	<0.001
Eastern	2.0 <i>1.013</i> (0.0–4.1)	2.3 <i>0.689</i> (0.9–3.7)	6.1 <i>1.764</i> (2.5–9.6)	0.053	0.56	88%	75%	72%	10%	20%	17%	2%	5%	9%	0%	0%	1%	<0.001
Lusaka	0.6 <i>0.303</i> (0.0–1.2)	0.9 <i>0.448</i> (0.03–1.8)	2.5 <i>0.792</i> (0.9–4.1)	0.030	0.45	93%	90%	85%	7%	8%	13%	0%	2%	2%	0%	0%	0%	0.020
Muchinga	2.0 <i>0.852</i> (0.2–3.7)	2.1 <i>0.799</i> (0.5–3.7)	3.8 <i>1.000</i> (1.8–5.8)	0.170	0.17	89%	81%	76%	7%	14%	17%	4%	5%	7%	0%	0%	0%	0.097
Southern	0.4 <i>0.294</i> (–0.2–1.0)	0.6 <i>0.241</i> (0.1–1.1)	2.5 <i>1.582</i> (–0.7–5.7)	0.194	0.56	97%	93%	86%	1%	6%	11%	1%	0%	3%	0%	0%	0%	<0.001

Exhibit 55: Task 5b – Reading Comprehension

Task 5b: Reading Comprehension (5 questions)																		
Province	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 questions correct)			% pre-emergent readers (1 question correct)			% emergent readers (2–3 questions correct)			% readers (4–5 questions correct)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
Central	1% 0.017 (0%–1%)	6% 0.050 (4%–8%)	8% 0.124 (2%–12%)	0.007	0.57	98%	92%	83%	2%	4%	12%	1%	4%	4%	0%	0%	1%	<0.001
Copperbelt	2% 0.043 (0%–4%)	4% 0.062 (1%–6%)	6% 0.099 (2%–10%)	0.029	0.44	95%	96%	90%	3%	1%	8%	2%	2%	1%	0%	0%	0%	0.001
Eastern	1% 0.033 (0%–2%)	6% 0.115 (2%–10%)	8% 0.127 (4%–14%)	0.009	0.53	97%	83%	83%	0%	9%	7%	1%	5%	8%	1%	3%	1%	<0.001
Lusaka	1% 0.053 (–1%–4%)	2% 0.058 (0%–4%)	2% 0.049 (1%–4%)	0.365	0.00	98%	94%	95%	0%	3%	3%	0%	1%	2%	1%	2%	0%	0.044
Muchinga	2% 0.065 (0%–6%)	4% 0.092 (0%–8%)	4% 0.044 (2%–6%)	0.577	0.15	94%	94%	93%	2%	2%	5%	4%	4%	2%	0%	0%	0%	0.117
Southern	0% 0.023 (0%–1%)	1% 0.017 (0%–2%)	4% 0.124 (–2%–8%)	0.238	0.56	98%	99%	93%	1%	1%	6%	1%	0%	1%	0%	0%	0%	0.005

Exhibit 56: Task 6 – Listening Comprehension: English¹³

Task 6: Listening Comprehension – English (5 questions)													
Province	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		p-value	Effect size	% scoring zero correct (0 questions correct)		% pre-emergent (1 question correct)		% emergent readers (2–3 questions correct)		% readers (4–5 questions correct)		p-value
	Midline	Endline			Midline	Endline	Midline	Endline	Midline	Endline	Midline	Endline	
Central	28% 0.215 (18%–36%)	28% 0.157 (22%–34%)	0.859	0.57	35%	31%	29%	32%	27%	29%	9%	7%	0.540
Copperbelt	26% 0.124 (22%–32%)	40% 0.368 (26%–56%)	0.077	0.32	35%	26%	35%	27%	24%	27%	6%	21%	<0.001
Eastern	8% 0.070 (6%–12%)	14% 0.144 (8%–20%)	0.098	0.21	76%	59%	12%	25%	11%	11%	2%	4%	<0.001
Lusaka	40% 0.167 (34%–46%)	36% 0.213 (28%–44%)	0.455	–0.09	45%	43%	8%	15%	31%	21%	16%	21%	0.003
Muchinga	8% 0.084 (4%–12%)	4% 0.072 (2%–8%)	0.196	–0.12	81%	81%	10%	15%	8%	3%	1%	0%	0.028
Southern	8% 0.123 (4%–14%)	12% 0.203 (4%–20%)	0.540	0.14	75%	67%	17%	24%	6%	6%	1%	3%	0.108

¹³ English listening comprehension was introduced into the Zambia EGRA at midline; reported p-values and effect sizes are for comparison of midline and endline results.

Exhibit 57: Task 7 – English Vocabulary

Task 7: English Vocabulary (20 words)					
	Mean percent questions correct Standard Error of Mean (95% confidence interval: lower, upper)	% first quartile	% second quartile	% third quartile	% fourth quartile
Province	Baseline	Baseline	Baseline	Baseline	Baseline
Central	40% 0.403 (36%–44%)	26%	24%	29%	22%
Copperbelt	33% 0.363 (30%–37%)	51%	22%	17%	11%
Eastern	n/a	n/a	n/a	n/a	n/a
Lusaka	n/a	n/a	n/a	n/a	n/a
Muchinga	50% 0.511 (44%–55%)	19%	9%	25%	47%
Southern	41% 0.431 (36%–45%)	28%	24%	27%	22%

¹⁴ English vocabulary was replaced by English listening comprehension after baseline. The information presented here is for historical purposes only.

EGRA Results by Language

Exhibit 58: Task 1 – Listening Comprehension: Zambian Language of Instruction ¹⁵

Task 1: Listening Comprehension: Zambian Language of Instruction (3 questions at baseline; 5 questions at midline and endline)																		
Language	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 questions correct)			% pre-emergent (1 question correct)			% emergent readers (2–3 questions correct)			% readers (4–5 questions correct)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
ChiTonga	90% <i>0.104</i> (83%–97%)	42% <i>0.135</i> (36%–46%)	46% <i>0.157</i> (38%–52%)	0.337	0.07	1%	31%	14%	2%	12%	17%	19%	31%	39%	77%	26%	29%	<0.001
CiNyanja	57% <i>0.142</i> (43%–63%)	62% <i>0.074</i> (60%–66%)	68% <i>0.066</i> (66%–72%)	0.006	0.16	19%	4%	2%	30%	6%	3%	22%	49%	41%	30%	41%	53%	<0.001
iCiBemba	60% <i>0.148</i> (50%–70%)	44% <i>0.099</i> (40%–48%)	50% <i>0.140</i> (44%–54%)	0.139	0.09	19%	14%	13%	26%	22%	19%	18%	43%	43%	38%	22%	24%	0.518

¹⁵ Baseline scores for Task 1 are not comparable to midline and endline scores due to limitations in the task at baseline; these scores are presented for informational purposes only. Reported p-values and effect sizes are for comparison of midline and endline scores.

Exhibit 59: Task 2 – Letter Sound Knowledge

Task 2: Letter Sound Knowledge (100 letters)																		
Language	Mean letters per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 letters per minute)			% pre-emergent readers (1– 5 letters per minute)			% emergent readers (16–40 letters per minute)			% readers (41+ letters per minute)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
ChiTonga	0.5 0.182 (0.1–0.8)	1.4 0.257 (0.8–1.9)	6.4 2.897 (0.5–12.3)	0.048	0.87	91%	77%	45%	8%	22%	45%	1%	1%	8%	0%	0%	2%	<0.001
CiNyanja	4.0 0.538 (2.9–5.0)	4.7 0.892 (2.9–6.4)	7.6 1.206 (5.2–10.0)	0.007	0.55	48%	50%	49%	44%	40%	33%	8%	9%	17%	0%	0%	1%	<0.001
iCiBemba	2.5 0.504 (1.5–3.5)	6.5 0.534 (5.5–7.6)	10.3 1.123 (8.1–12.5)	<0.001	0.95	91%	77%	45%	8%	22%	45%	1%	1%	8%	0%	0%	2%	<0.001

Exhibit 60: Task 3 – Non-word Reading

Task 3: Non-word Reading (50 invented words)																		
Language	Mean words per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 words per minute)			% pre-emergent readers (1–14 words per minute)			% emergent readers (15–29 words per minute)			% readers (30+ words per minute)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
ChiTonga	0.3 0.215 (–0.1–0.7)	0.6 0.244 (0.1–1.1)	2.4 1.494 (–0.7–5.4)	0.183	0.66	96%	92%	84%	3%	7%	11%	1%	1%	6%	0%	0%	0%	<0.001
CiNyanja	1.0 0.334 (0.3–1.7)	1.5 0.353 (0.8–2.2)	3.5 0.873 (1.8–5.3)	0.009	0.56	85%	82%	77%	13%	13%	15%	2%	4%	7%	0%	0%	1%	<0.001
iCiBemba	0.9 0.301 (0.3–1.5)	2.4 0.404 (1.6–3.2)	4.7 0.844 (3.0–6.4)	<0.001	0.58	90%	78%	68%	7%	16%	18%	2%	5%	12%	0%	1%	1%	<0.001

Exhibit 61: Task 4 – Orientation to Print¹⁶

Task 4: Orientation to Print (3 questions)													
Language	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		p-value	Effect size	% scoring zero correct (0 questions correct)		% pre-emergent readers (1 question correct)		% emergent readers (2 question correct)		% readers (3 question correct)		p-value
	Midline	Endline			Mid	End	Mid	End	Mid	End	Mid	End	
ChiTonga	83% 0.891 (77%–87%)	93% 0.060 (90%–97%)	0.005	0.20	12%	2%	4%	4%	12%	12%	71%	82%	<0.001
CiNyanja	87% 0.049 (83%–90%)	93% 0.052 (90%–97%)	0.003	0.17	12%	6%	2%	3%	7%	4%	79%	86%	0.001
iCiBemba	73% 0.082 (70%–80%)	90% 0.035 (87%–90%)	<0.001	0.22	15%	6%	7%	5%	10%	8%	68%	82%	<0.001

¹⁶ Task 4 was only introduced during the final period of data collection at baseline; consequently, data are not available for Eastern and Lusaka provinces and sample sizes for the other provinces are too small to enable statistical comparison. Reported p-values and effect sizes are for comparison of midline and endline scores.

Exhibit 62: Task 5a – Oral Passage Reading

Task 5a: Oral Passage Reading (40–68 words, depending on EGRA language version and evaluation phase)																		
Language	Mean words per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 words per minute)			% pre-emergent readers (1–19 words per minute)			% emergent readers (20–44 words per minute)			% readers (45+ words per minute)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
ChiTonga	0.4 <i>0.294</i> (–0.2–1.0)	0.6 <i>0.241</i> (0.1–1.1)	2.5 <i>1.582</i> (–0.7–5.7)	0.194	0.56	97%	93%	86%	1%	6%	11%	1%	0%	3%	0%	0%	0%	<0.001
CiNyanja	1.1 <i>0.400</i> (0.3–1.9)	1.5 <i>0.409</i> (0.7–2.3)	4.7 <i>0.409</i> (2.2–7.2)	0.007	0.59	90%	82%	78%	9%	14%	15%	1%	4%	6%	0%	0%	1%	<0.001
iCiBemba	1.1 <i>0.412</i> (0.3–1.9)	2.2 <i>0.385</i> (1.4–2.9)	5.0 <i>0.982</i> (3.1–7.0)	<0.001	0.51	92%	79%	72%	5%	18%	19%	3%	3%	9%	0%	0%	0%	<0.001

Exhibit 63: Task 5b – Reading Comprehension

Task 5b: Reading Comprehension (5 questions)																		
Language	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			p-value	Effect size	% scoring zero correct (0 questions correct)			% pre-emergent readers (1 question correct)			% emergent readers (2–3 questions correct)			% readers (4–5 questions correct)			p-value
	Baseline	Midline	Endline			Base	Mid	End	Base	Mid	End	Base	Mid	End	Base	Mid	End	
ChiTonga	0% <i>0.023</i> (0%–1%)	1% <i>0.017</i> (0%–2%)	4% <i>0.124</i> (–2%–8%)	0.238	0.56	98%	99%	93%	1%	1%	6%	1%	0%	1%	0%	0%	0%	0.005
CiNyanja	1% <i>0.037</i> (0%–2%)	4% <i>0.062</i> (1%–6%)	6% <i>0.088</i> (2%–10%)	0.013	0.46	98%	88%	89%	0%	6%	5%	1%	3%	5%	1%	2%	1%	<0.001
iCiBemba	1% <i>0.022</i> (0%–2%)	4% <i>0.041</i> (2%–6%)	6% <i>0.064</i> (4%–8%)	<0.001	0.44	96%	94%	89%	2%	3%	9%	2%	3%	3%	0%	0%	0%	<0.001

Exhibit 64: Task 6 – Listening Comprehension: English¹⁷

Task 6: Listening Comprehension – English (5 questions)													
Language	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		p-value	Effect size	% scoring zero correct (0 questions correct)		% pre-emergent (1 question correct)		% emergent readers (2–3 questions correct)		% readers (4–5 questions correct)		p-value
	Midline	Endline			Midline	Endline	Midline	Endline	Midline	Endline	Midline	Endline	
ChiTonga	8% 0.123 (4%–14%)	12% 0.203 (4%–20%)	0.540	0.14	75%	67%	17%	24%	6%	6%	1%	3%	0.108
CiNyanja	26% 0.186 (20%–34%)	22% 0.148 (16%–28%)	0.457	–0.09	60%	51%	10%	20%	21%	16%	9%	13%	<0.001
iCiBemba	22% 0.100 (18%–26%)	28% 0.203 (22%–38%)	0.153	0.14	48%	46%	26%	25%	20%	25%	6%	9%	0.059

Exhibit 65: Task 7 – English Vocabulary¹⁸

Task 7: English Vocabulary (20 words)					
Language	Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)	% first quartile	% second quartile	% third quartile	% fourth quartile
	Baseline	Baseline	Baseline	Baseline	Baseline
ChiTonga	41% 0.431 (36%–45%)	28%	24%	27%	22%
CiNyanja	n/a	n/a	n/a	n/a	n/a

¹⁷ English listening comprehension was introduced into the Zambia EGRA at midline; reported p-values and effect sizes are for comparison of midline and endline results.

¹⁸ English vocabulary was replaced by English listening comprehension after baseline. The information presented here is for historical purposes only.

iCiBemba	39% 0.252 (36%–41%)	34%	19%	23%	24%
----------	---------------------------	-----	-----	-----	-----

EGRA Results by Sex

Exhibit 66: Task 1 – Listening Comprehension: Zambian Language of Instruction

Province	Sex	Task 1: Listening Comprehension: Zambian Language of Instruction (3 questions at baseline; 5 questions at midline and endline)														
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)														
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Central	F	50% 0.219 (33%–63%)	46% 0.209 (38%–56%)	56% 1.915 (38%–58%)	29%	30%	23%	18%	17%	14%	44%	26%	14%	18%	39%	29%
	M	53% 0.194 (40%–67%)	52% 0.147 (46%–58%)	48% 2.471 (50%–64%)	22%	25%	30%	23%	7%	12%	56%	25%	5%	21%	38%	36%
	P	0.325	0.224	0.018	0.478				0.041				0.021			
Copperbelt	F	67% 0.275 (47%–87%)	40% 0.175 (32%–48%)	52% 0.299 (40%–64%)	7%	36%	13%	44%	17%	26%	40%	17%	13%	19%	44%	24%
	M	70% 0.248 (53%–87%)	46% 0.150 (40%–52%)	52% 0.213 (42%–62%)	3%	35%	10%	52%	11%	27%	41%	21%	9%	12%	53%	26%
	P	0.367	0.012	0.799	0.554				0.498				0.206			
Eastern	F	60% 0.187 (47%–73%)	58% 0.131 (54%–64%)	68% 0.084 (64%–72%)	12%	33%	28%	28%	7%	9%	45%	40%	4%	4%	37%	55%
	M	57% 0.176 (47%–70%)	64% 0.175 (56%–70%)	70% 0.142 (64%–76%)	9%	39%	17%	35%	4%	7%	49%	40%	4%	3%	31%	63%
	P	0.666	0.129	0.731	0.136				0.645				0.536			
Lusaka	F	57% 0.197 (43%–67%)	62% 0.172 (56%–79%)	66% 0.118 (64%–70%)	31%	20%	21%	28%	4%	5%	48%	43%	1%	3%	53%	42%
	M	53% 0.210 (37%–67%)	66% 0.058 (62%–68%)	70% 0.134 (60%–76%)	26%	27%	20%	27%	1%	3%	55%	41%	2%	3%	45%	49%

Province	Sex	Task 1: Listening Comprehension: Zambian Language of Instruction (3 questions at baseline; 5 questions at midline and endline)														
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)														
					Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
	P	0.228	0.445	0.021	0.669				0.309				0.623			
Muchinga	F	53% 0.249 (33%–70%)	38% 0.214 (30%–46%)	36% 0.195 (28%–44%)	35%	10%	19%	35%	20%	28%	33%	19%	18%	26%	47%	9%
	M	70% 0.303 (50%–90%)	44% 0.227 (36%–54%)	38% 0.157 (32%–44%)	27%	3%	10%	60%	11%	24%	42%	22%	20%	20%	42%	18%
	P	0.007	0.203	0.834	0.040				0.191				0.167			
Southern	F	90% 0.106 (83%–100%)	42% 0.238 (32%–52%)	42% 0.202 (34%–52%)	3%	1%	20%	76%	31%	14%	29%	26%	18%	15%	41%	26%
	M	90% 0.105 (83%–100%)	40% 0.159 (34%–46%)	48% 0.129 (44%–54%)	0%	3%	19%	78%	31%	10%	34%	25%	10%	19%	37%	34%
	P	0.943	0.706	0.049	0.519				0.673				0.111			

Exhibit 67: Task 2 – Letter Sound Knowledge

Province	Sex	Task 2: Letter Sound Knowledge (100 letters)														
		Mean letters per minute Standard Error of Mean (95% confidence interval: lower, upper)			Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Central	F	2.4 0.603 (1.2–3.6)	8.7 0.917 (6.9–10.6)	11.6 1.233 (9.1–14.1)	71%	24%	5%	0%	33%	53%	13%	1%	22%	50%	28%	0%
	M	2.1 0.551 (1.0–3.2)	9.7 1.291 (7.0–12.3)	10.7 1.106 (8.5–12.9)	71%	23%	6%	0%	22%	58%	20%	0%	27%	49%	24%	0%
	P	0.147	0.423	0.366	0.972				0.113				0.540			
Copperbelt	F	2.2 0.730 (0.7–3.7)	5.7 0.640 (4.4–7.0)	11.5 2.525 (6.3–16.6)	73%	25%	2%	0%	39%	56%	5%	0%	31%	51%	17%	1%
	M	2.1 1.119 (−0.2–4.4)	4.6 0.564 (3.4–5.7)	9.1 2.117 (4.8–13.5)	73%	23%	4%	0%	41%	55%	4%	0%	31%	45%	24%	0%
	P	0.876	0.022	0.030	0.562				0.845				0.245			
Eastern	F	2.7 0.915 (0.9–4.6)	5.7 0.807 (4.0–7.3)	9.0 1.261 (6.4–11.5)	61%	31%	8%	0%	42%	46%	12%	0%	33%	45%	22%	0%
	M	4.3 1.473 (1.3–7.3)	5.3 0.786 (3.7–6.9)	9.4 2.072 (5.2–13.5)	55%	32%	11%	1%	34%	58%	9%	0%	31%	39%	29%	1%
	P	0.036	0.652	0.794	0.505				0.133				0.375			
Lusaka	F	3.7 0.667 (2.4–5.1)	4.4 1.898 (0.5–8.2)	4.9 1.996 (0.9–8.9)	36%	60%	5%	0%	59%	31%	10%	1%	67%	24%	6%	3%
	M	4.7 0.762 (3.1–6.2)	3.6 1.150 (1.3–6.0)	5.5 1.327 (2.8–8.2)	33%	60%	8%	0%	67%	26%	7%	0%	65%	23%	11%	1%
	P	0.294	0.539	0.569	0.595				0.365				0.154			

Province	Sex	Task 2: Letter Sound Knowledge (100 letters)														
		Mean letters per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Muchinga	F	4.2 1.450 (1.3–7.2)	6.1 1.645 (2.8–9.5)	9.2 1.972 (5.2–13.2)	61%	32%	6%	0%	33%	57%	11%	0%	43%	32%	24%	0%
	M	4.0 1.305 (1.3–6.6)	7.0 0.965 (5.0–8.9)	7.4 0.962 (5.4–9.3)	48%	45%	5%	2%	31%	58%	11%	0%	40%	48%	12%	0%
	P	0.497	0.471	0.253	0.339				0.974				0.005			
Southern	F	0.6 0.318 (0.0–1.2)	1.9 0.465 (0.9–2.8)	6.5 3.327 (–0.3–13.2)	88%	12%	0%	0%	70%	29%	1%	0%	48%	43%	7%	3%
	M	0.3 0.224 (–0.1–0.8)	0.8 0.219 (0.4–1.3)	6.3 2.402 (1.5–11.2)	94%	4%	1%	0%	83%	15%	1%	1%	42%	47%	9%	0%
	P	0.486	0.052	0.887	0.162				0.034				0.583			

Exhibit 68: Task 3 – Non-word Reading

Province	Sex	Task 3: Non-word Reading (50 invented words)														
		Mean words per minute Standard Error of Mean (95% confidence interval: lower, upper)			Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Central	F	0.6 0.414 (−0.2–1.4)	1.8 0.477 (0.8–2.8)	5.4 1.443 (2.4–8.3)	93%	6%	1%	0%	75%	21%	4%	0%	65%	17%	15%	3%
	M	0.3 0.191 (−0.1–0.7)	3.8 0.967 (1.9–5.8)	5.1 1.490 (2.1–8.1)	98%	1%	1%	0%	67%	22%	10%	1%	62%	18%	19%	1%
	P	0.185	0.020	0.740	0.216				0.145				0.528			
Copperbelt	F	1.1 0.451 (0.2–2.0)	1.0 0.309 (0.4–1.6)	5.1 1.368 (2.3–7.9)	90%	8%	2%	0%	85%	14%	1%	0%	69%	20%	11%	0%
	M	0.9 0.764 (−0.6–2.5)	3.5 1.131 (1.2–5.8)	4.4 1.643 (1.0–7.7)	89%	9%	2%	0%	85%	9%	4%	3%	71%	19%	8%	2%
	P	0.776	0.019	0.382	0.926				0.040				0.383			
Eastern	F	1.1 0.719 (−0.3–2.6)	2.1 0.654 (0.8–3.4)	4.4 1.255 (1.9–6.9)	87%	13%	0%	0%	80%	14%	6%	0%	71%	20%	9%	1%
	M	2.1 0.917 (0.2–3.9)	2.1 0.520 (1.0–3.1)	4.4 1.414 (1.5–7.2)	79%	17%	4%	0%	71%	23%	6%	0%	69%	16%	12%	3%
	P	0.048	0.940	0.994	0.026				0.155				0.347			
Lusaka	F	0.8 0.559 (−0.3–2.0)	1.0 0.489 (0.0–2.0)	1.9 0.779 (0.3–3.5)	92%	7%	1%	0%	89%	10%	1%	0%	84%	14%	3%	0%
	M	0.7 0.263 (0.1–1.2)	1.3 0.548 (0.2–2.4)	2.4 0.709 (1.0–3.9)	80%	17%	3%	0%	89%	7%	4%	0%	85%	11%	4%	0%
	P	0.809	0.430	0.472	0.018				0.374				0.708			
Muchinga	F	0.7 0.280 (0.1–1.3)	1.8 0.884 (0.0–1.6)	3.3 0.936 (1.4–5.2)	92%	6%	2%	0%	82%	13%	4%	1%	75%	13%	12%	1%
	M	3.2 1.280 (0.6–5.8)	2.9 0.653 (1.6–4.3)	3.8 0.912 (2.0–5.7)	77%	17%	3%	0%	69%	22%	9%	0%	68%	22%	8%	2%
	P	0.025	0.110	0.403	0.109				0.072				0.130			

Province	Sex	Task 3: Non-word Reading (50 invented words)														
		Mean words per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Southern	F	0.4 0.424 (−0.4–1.3)	0.6 0.400 (−0.2–1.4)	2.6 1.825 (−1.1–6.3)	95%	4%	1%	0%	93%	7%	1%	0%	86%	6%	7%	0%
	M	0.2 0.183 (−0.2–0.6)	0.7 0.312 (0.0–1.3)	2.1 1.086 (−0.1–4.3)	97%	3%	0%	0%	92%	7%	1%	0%	81%	16%	4%	0%
	P	0.604	0.834	0.492	0.601				0.995				0.013			

Exhibit 69: Task 4 – Orientation to Print

Province	Sex	Task 4: Orientation to Print (3 questions)									
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		Midline				Endline			
		Midline	Endline	0	1	2	3	0	1	2	3
Central	F	93% 0.063 (90%–97%)	90% 0.060 (87%–97%)	5%	2%	5%	89%	4%	4%	6%	86%
	M	93% 0.063 (90%–100%)	90% 0.052 (87%–97%)	1%	3%	6%	89%	5%	3%	8%	85%
	P	0.529	0.941	0.288				0.826			
Copperbelt	F	57% 0.114 (50%–63%)	87% 0.054 (83%–90%)	35%	11%	10%	44%	4%	7%	6%	83%
	M	63% 0.141 (53%–73%)	93% 0.092 (87%–100%)	27%	15%	12%	47%	3%	3%	3%	91%
	P	0.090	0.046	0.404				0.267			
Eastern	F	77% 0.105 (70%–83%)	90% 0.104 (83%–97%)	20%	3%	8%	69%	8%	4%	1%	87%
	M	77% 0.119 (67%–83%)	93% 0.078 (87%–97%)	17%	3%	15%	65%	8%	3%	2%	88%
	P	0.878	0.602	0.324				0.884			
Lusaka	F	93% 0.093 (87%–100%)	93% 0.079 (87%–97%)	4%	1%	2%	94%	5%	4%	5%	85%
	M	90% 0.128 (83%–100%)	93% 0.068 (90%–100%)	6%	1%	3%	90%	4%	2%	7%	87%
	P	0.676	0.377	0.608				0.605			

Province	Sex	Task 4: Orientation to Print (3 questions)									
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		Midline				Endline			
		Midline	Endline	0	1	2	3	0	1	2	3
Muchinga	F	70% 0.126 (70%–87%)	83% 0.128 (70%–87%)	13%	6%	17%	64%	8%	7%	18%	67%
	M	87% 0.077 (87%–100%)	83% 0.072 (87%–100%)	5%	3%	8%	84%	9%	5%	8%	77%
	p	0.010	0.662	0.008				0.077			
Southern	F	80% 0.094 (73%–90%)	90% 0.088 (73%–87%)	13%	6%	12%	69%	4%	4%	12%	81%
	M	87% 0.107 (80%–90%)	97% 0.052 (80%–93%)	11%	2%	13%	74%	1%	4%	11%	84%
	p	0.043	0.104	0.345				0.390			

Exhibit 70: Task 5a – Oral Passage Reading

Province	Sex	Task 5a: Oral Passage Reading (40– 68 words, depending on EGRA language version and evaluation phase)														
		Mean words per minute Standard Error of Mean (95% confidence interval: lower, upper)														
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Central	F	0.8 0.417 (–0.1–1.5)	2.1 0.580 (0.93–3.2)	5.3 1.80 (1.5–8.8)	94%	6%	0%	0%	75%	24%	1%	0%	67%	18%	14%	1%
	M	0.3 0.297 (–0.3–0.9)	4.1 0.977 (2.0–6.0)	5.2 1.64 (1.8–8.5)	98%	1%	1%	0%	68%	29%	3%	0%	67%	21%	11%	1%
	P	0.171	0.027	0.990	0.139				0.191				0.719			
Copperbelt	F	1.4 0.641 (0.1–2.7)	1.4 0.384 (0.7–2.2)	5.2 1.48 (2.1–8.2)	92%	5%	3%	0%	83%	16%	1%	0%	71%	21%	7%	0%
	M	1.3 1.176 (–1.1–3.7)	4.2 1.379 (1.4–7.0)	3.7 1.396 (0.9–6.6)	90%	4%	6%	0%	83%	11%	6%	0%	73%	23%	4%	0%
	P	0.904	0.039	0.070	0.628				0.015				0.604			
Eastern	F	1.3 0.977 (–0.7–3.3)	2.6 0.947 (0.7–4.5)	6.1 1.86 (2.4–9.8)	90%	10%	0%	0%	74%	22%	4%	0%	73%	17%	9%	1%
	M	2.7 1.153 (0.4–5.0)	2.8 0.904 (1.0–4.6)	6.0 1.838 (2.3–9.7)	84%	12%	4%	0%	75%	19%	6%	0%	72%	17%	10%	2%
	P	0.045	0.829	0.896	0.040				0.599				0.808			
Lusaka	F	0.6 0.529 (–0.4–1.7)	1.1 0.556 (–0.1–2.2)	2.1 0.915 (0.2–3.9)	96%	4%	0%	0%	88%	11%	1%	0%	85%	13%	2%	0%
	M	0.5 0.191 (0.1–0.9)	1.2 0.583 (0.0–2.3)	2.8 0.953 (0.9–4.8)	88%	12%	0%	0%	92%	5%	3%	0%	85%	13%	3%	0%
	P	0.837	0.822	0.470	0.024				0.087				0.940			
Muchinga	F	0.5 0.334 (–0.2–1.1)	2.2 1.306 (–0.5–4.8)	3.3 0.936 (1.4–5.2)	95%	3%	2%	0%	85%	11%	3%	0%	77%	17%	6%	0%
	M	3.4 1.451 (0.5–6.4)	3.3 0.909 (1.5–5.2)	3.6 0.977 (1.7–5.6)	83%	10%	7%	0%	75%	18%	7%	0%	76%	16%	7%	1%
	P	0.026	0.186	0.595	1.105				0.156				0.744			

Province	Sex	Task 5a: Oral Passage Reading (40– 68 words, depending on EGRA language version and evaluation phase)														
		Mean words per minute <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)			Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Southern	F	0.6 0.613 (−0.6–1.9)	0.6 0.395 (−0.2–1.4)	3.0 2.082 (−1.2–7.2)	96%	1%	3%	0%	92%	7%	1%	0%	88%	8%	4%	0%
	M	0.1 0.106 (−0.1–0.3)	0.5 0.228 (0.0–1.0)	2.5 1.374 (−0.3–5.3)	99%	1%	0%	0%	94%	6%	0%	0%	84%	14%	2%	0%
	P	0.408	0.819	0.555	0.403				0.562				1.187			

Exhibit 7I: Task 5b – Reading Comprehension

Province	Sex	Task 5b: Reading Comprehension (5 questions)														
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)														
		Baseline	Midline	Endline	Baseline				Midline				Endline			
					0	1	2	3	0	1	2	3	0	1	2	3
Central	F	1% 0.035 (0%–0%)	4% 0.054 (1%–6%)	8% 0.149 (2%–14%)	95%	4%	1%	0%	96%	1%	3%	0%	83%	13%	4%	0%
	M	0% 0 (0%–0%)	8% 0.072 (4%–10%)	10% 0.164 (2%–16%)												
	P				100%	0%	0%	0%	88%	7%	5%	0%	84%	11%	4%	1%
		0.120	0.003	0.769	0.112				0.034				0.755			
Copperbelt	F	2% 0.034 (0%–2%)	1% 0.027 (0%–2%)	8% 0.125 (4%–14%)	96%	2%	2%	0%	99%	1%	1%	0%	89%	10%	1%	0%
	M	2% 0.072 (–1%–4%)	6% 0.123 (2%–12%)	6% 0.139 (1%–12%)	93%	3%	35	0%	93%	2%	4%	1%	92%	7%	1%	0%
	P	0.979	0.023	0.383	0.655				0.101				0.502			
Eastern	F	1% 0.023 (–1%–2%)	6% 0.133 (1%–12%)	12% 0.172 (4%–14%)	99%	0%	1%	1%	83%	10%	4%	3%	86%	6%	7%	1%
	M	2% 0.059 (0%–4%)	6% 0.106 (2%–10%)	12% 0.173 (6%–20%)	97%	1%	2%	0%	81%	8%	7%	3%	80%	8%	10%	2%
	P	0.356	0.848	0.729	0.327				0.729				0.584			
Lusaka	F	2% 0.099 (–2%–2%)	2% 0.044 (0%–4%)	4% 0.089 (0%–8%)	98%	1%	1%	0%	92%	4%	3%	1%	95%	4%	1%	0%
	M	1% 0.038 (–1%–6%)	2% 0.075 (–1%–6%)	6% 0.097 (2%–10%)	98%	0%	0%	2%	95%	2%	0%	3%	95%	3%	2%	0%
	P	0.533	0.892	0.342	0.251				0.076				0.728			

Province	Sex	Task 5b: Reading Comprehension (5 questions)														
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)														
					Baseline				Midline				Endline			
		Baseline	Midline	Endline	0	1	2	3	0	1	2	3	0	1	2	3
Muchinga	F	1% <i>0.023</i> (0%–1%)	2% <i>0.110</i> (–2%–8%)	4% <i>0.060</i> (2%–8%)												
	M	6% <i>0.119</i> (0%–10%)	4% <i>0.100</i> (1%–8%)	6% <i>0.074</i> (2%–8%)												
	P															
		0.040	0.151	0.565	0.123				0.139				0.876			
Southern	F	1% <i>0.047</i> (1%–1%)	1% <i>0.032</i> (0%–2%)	6% <i>0.175</i> (–2%–12%)												
	M	0% <i>0</i> (0%–1%)	1% <i>0.023</i> (0%–2%)	4% <i>0.115</i> (–1%–8%)												
	P															
		0.301	0.954	0.202	0.254				0.620				0.903			

Exhibit 72: Task 6 – Listening Comprehension: English

Province	Sex	Task 6: Listening Comprehension – English (5 questions)									
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		Midline				Endline			
		Midline	Endline	0	1	2	3	0	1	2	3
Central	F	26% 0.229 (20%–34%)	26% 0.168 (20%–34%)	38%	31%	22%	9%	31%	34%	29%	7%
	M	28% 0.237 (16%–38%)	30% 0.198 (22%–38%)	33%	27%	31%	10%	32%	31%	30%	8%
	p	0.364	0.600	0.436				0.937			
Copperbelt	F	26% 0.161 (20%–32%)	44% 0.402 (28%–60%)	36%	34%	23%	19%	25%	23%	30%	22%
	M	28% 0.136 (22%–32%)	36% 0.340 (22%–50%)	33%	36%	25%	6%	27%	30%	23%	20%
	p	0.746	0.132	0.928				0.402			
Eastern	F	6% 0.067 (4%–10%)	16% 0.157 (10%–22%)	82%	8%	11%	0%	62%	25%	11%	3%
	M	12% 0.101 (8%–16%)	14% 0.177 (6%–20%)	70%	16%	10%	3%	57%	26%	12%	5%
	p	0.015	0.433	0.011				0.796			
Lusaka	F	34% 0.233 (24%–44%)	36% 0.237 (26%–46%)	48%	6%	31%	14%	42%	16%	22%	20%
	M	46% 0.210 (38%–54%)	36% 0.222 (26%–44%)	41%	10%	31%	18%	43%	15%	19%	23%
	p	0.056	0.714	0.445				0.900			
Muchinga	F	4% 0.113 (0%–10%)	4% 0.076 (2%–8%)	88%	7%	5%	0%	80%	18%	1%	1%
	M	12% 0.094 (8%–14%)	6% 0.089 (2%–8%)	74%	13%	11%	2%	82%	12%	5%	0%
	p	0.009	0.557	0.037				0.041			

Province	Sex	Task 6: Listening Comprehension – English (5 questions)									
		Mean percent questions correct <i>Standard Error of Mean</i> (95% confidence interval: lower, upper)		Midline				Endline			
		Midline	Endline	0	1	2	3	0	1	2	3
Southern	F	10% 0.075 (2%–16%)	10% 0.212 (2%–8%)	77%	15%	7%	1%	69%	23%	6%	2%
	M	8% 0.231 (4%–12%)	14% 0.197 (6%–22%)	73%	20%	5%	1%	64%	25%	7%	4%
	p	0.774	0.031	0.599				0.688			

ANNEX 2. CLASSROOM OBSERVATION SUMMARY TABLES

Exhibit 73: Percentage of Lessons during which Criteria Were Fulfilled at least Once, Midline vs. Endline

Domain	Criterion	Definition	Midline (n=102)	Endline (n=105)	Difference	Significance (p-value) *
I. Orientation to Print	1	Orientation to print	n/a	82%	n/a	n/a
II. Phonemic Awareness and Phonics	2	Teacher teaches letter sounds	88%	79%	-9%	0.074
II. Phonemic Awareness and Phonics	3	Teacher demonstrates phonemic awareness	50%	31%	-19%	0.007
II. Phonemic Awareness and Phonics	4	Learners demonstrate phonemic awareness	58%	42%	-16%	0.022
II. Phonemic Awareness and Phonics	5	Teacher teaches syllables	81%	72%	-9%	0.125
II. Phonemic Awareness and Phonics	6	Teacher teaches coding and decoding	74%	60%	-14%	0.039
II. Phonemic Awareness and Phonics	7	Teacher sounds out words	51%	70%	19%	0.006
II. Phonemic Awareness and Phonics	8	Learners sound out words	47%	78%	31%	<0.001
II. Phonemic Awareness and Phonics	9	Teacher dictates material for learners to write	n/a	25%	n/a	n/a
III. Passage/Story Reading	10	Teacher models reading	53%	65%	12%	0.084
III. Passage/Story Reading	11	Learners read connected text	32%	65%	33%	<0.001
IV. Reading Comprehension	12	Learners asked reading comprehension questions	13%	37%	24%	<0.001
IV. Reading Comprehension	13	Learners asked prediction questions while reading	5%	12%	7%	0.056
IV. Reading Comprehension	14	Teacher covers vocabulary in learners' reading material	12%	31%	19%	0.001
V. Oral Language	15	Teacher checks listening comprehension while reading	34%	39%	5%	0.48
V. Oral Language	16	Teacher asks prediction questions while reading	9%	24%	15%	0.004
V. Oral Language	17	Teaching uses dialogue for teaching listening skills	27%	51%	24%	<0.001
VI. Writing	18	Learners practice forming letters	n/a	15%	n/a	n/a
VI. Writing	19	Learners copy words or letters	75%	63%	-0.12	0.071
VI. Writing	20	Learners write or draw freely	23%	32%	0.09	0.113

* Chi-square test for significance of difference of samples

Exhibit 74: Percentage of Lessons during which Lesson Delivery Criteria Were Fulfilled at least Once at Endline

Domain	Criterion	Definition	Endline (n=105)
Part B: Lesson Delivery	21	Teacher presenting to whole class	90%
Part B: Lesson Delivery	22	Call and response	99%
Part B: Lesson Delivery	23	Learner demonstrating for class	97%
Part B: Lesson Delivery	24	Learners work in small groups or alone	83%
Part B: Lesson Delivery	25	Teacher checks learners' work	78%
Part B: Lesson Delivery	26	Teacher not interacting with learners	75%

Exhibit 75: Percentage of Lessons during which Criteria Were Fulfilled at least Once for Criteria Used at Midline Only

Domain	Definition	Midline (n=102)
I. Orientation to Print	Teacher demonstrates orientation to print	83%
I. Orientation to Print	Teacher watches learners demonstrate orientation to print	17%
I. Orientation to Print	Teacher instructs orientation to print orally	24%
I. Orientation to Print	Teacher instructs orientation to print visually	23%
VI. Writing	Teacher practices letters' strokes and curves with learners	1%
VI. Writing	Learners write strokes and curves of letters	12%

Exhibit 76: Number of 3-Minute Intervals Criteria Were Performed at Midline and Endline

Domain	Criterion	Definition	Midline (n=102)				Endline (n=105)				Comparison: Midline to Endline	
			Mean # intervals	Standard Deviation	Min	Max	Mean # intervals	Standard Deviation	Min	Max	Change in mean # intervals	Significance (p-value)*

Domain	Criterion	Definition	Midline (n=102)				Endline (n=105)				Comparison: Midline to Endline	
			Mean # intervals	Standard Deviation	Min	Max	Mean # intervals	Standard Deviation	Min	Max	Change in mean # intervals	Significance (p-value)*
I. Orientation to Print	1	Orientation to print	n/a	n/a	n/a	n/a	6.3	5.813	0	20	n/a	n/a
II. Phonemic Awareness and Phonics	2	Teacher teaches letter sounds	5.4	4.003	0	19	3.2	3.274	0	16	-2.2	<0.001
II. Phonemic Awareness and Phonics	3	Teacher demonstrates phonemic awareness	2.3	3.274	0	15	0.9	1.557	0	7	-1.5	<0.001
II. Phonemic Awareness and Phonics	4	Learners demonstrate phonemic awareness	2.7	3.381	0	14	1.2	1.865	0	9	-1.5	<0.001
II. Phonemic Awareness and Phonics	5	Teacher teaches syllables	5.5	4.673	0	18	3.6	3.834	0	17	-1.9	0.002
II. Phonemic Awareness and Phonics	6	Teacher teaches coding and decoding	4.0	4.167	0	15	2.7	3.203	0	14	-1.3	0.012
II. Phonemic Awareness and Phonics	7	Teacher sounds out words	1.6	2.481	0	12	2.7	2.772	0	11	1.0	0.0046
II. Phonemic Awareness and Phonics	8	Learners sound out words	1.5	2.228	0	9	3.5	3.291	0	15	2.0	<0.001
II. Phonemic Awareness and Phonics	9	Teacher dictates material for learners to write	n/a	n/a	n/a	n/a	0.8	1.890	0	13	n/a	n/a
III. Passage/Story Reading	10	Teacher models reading	1.4	2.020	0	12	1.9	2.233	0	11	0.5	0.069
III. Passage/Story Reading	11	Learners read connected text	1.5	3.296	0	18	2.4	2.642	0	13	0.9	0.031
IV. Reading Comprehension	12	Learners asked reading comprehension	0.4	1.484	0	11	0.9	1.559	0	8	0.5	0.023

Domain	Criterion	Definition	Midline (n=102)				Endline (n=105)				Comparison: Midline to Endline	
			Mean # intervals	Standard Deviation	Min	Max	Mean # intervals	Standard Deviation	Min	Max	Change in mean # intervals	Significance (p-value)*
		questions										
IV. Reading Comprehension	13	Learners asked prediction questions while reading	0.1	0.323	0	2	0.3	0.835	0	4	0.2	0.025
IV. Reading Comprehension	14	Teacher covers vocabulary in learners' reading material	0.3	1.134	0	9	0.7	1.411	0	7	0.4	0.043
V. Oral Language	15	Teacher checks listening comprehension while reading	0.5	0.875	0	4	0.9	1.520	0	7	0.4	0.026
V. Oral Language	16	Teacher asks prediction questions while reading	0.1	0.438	0	2	0.4	0.909	0	4	0.3	0.002
V. Oral Language	17	Teaching uses dialogue for teaching listening skills	0.6	1.424	0	11	2.0	2.813	0	17	1.4	<0.001
VI. Writing	18	Learners practice forming letters	n/a	n/a	n/a	n/a	0.3	0.796	0	4	n/a	n/a
VI. Writing	19	Learners copy words or letters	5.7	4.793	0	16	2.4	2.837	0	12	-3.3	<0.001
VI. Writing	20	Learners write or draw freely	1.4	3.164	0	15	1.0	1.944	0	10	-0.3	0.378
Part B: Lesson Delivery	B21	Teacher presents to whole class	n/a	n/a	n/a	n/a	8.1	5.421	0	20	n/a	n/a
Part B: Lesson Delivery	B22	Call and response	n/a	n/a	n/a	n/a	8.3	4.501	0	20	n/a	n/a
Part B: Lesson Delivery	B23	Learner demonstrating for class	n/a	n/a	n/a	n/a	7.4	4.292	0	18	n/a	n/a
Part B: Lesson Delivery	B24	Learners work in small groups or alone	n/a	n/a	n/a	n/a	5.4	4.031	0	16	n/a	n/a
Part B: Lesson Delivery	B25	Teacher checks learners' work	n/a	n/a	n/a	n/a	4.3	3.778	0	15	n/a	n/a

Domain	Criterion	Definition	Midline (n=102)				Endline (n=105)				Comparison: Midline to Endline	
			Mean # intervals	Standard Deviation	Min	Max	Mean # intervals	Standard Deviation	Min	Max	Change in mean # intervals	Significance (p-value)*
Part B: Lesson Delivery	B26	Teacher does not interact with learners	n/a	n/a	n/a	n/a	2.7	3.102	0	16	n/a	n/a
Midline only	Midline 1	Teacher demonstrates orientation to print	6.1	4.910	0	20	n/a	n/a	n/a	n/a	n/a	n/a
Midline only	Midline 2	Teacher watches learners demonstrate orientation to print	0.4	1.309	0	9	n/a	n/a	n/a	n/a	n/a	n/a
Midline only	Midline 3	Teacher instructs orientation to print orally	0.5	1.355	0	9	n/a	n/a	n/a	n/a	n/a	n/a
Midline only	Midline 4	Teacher instructs orientation to print visually	1.2	2.663	0	10	n/a	n/a	n/a	n/a	n/a	n/a
Midline only	Midline 18	Teacher practices letters' strokes and curves with learners	0.0	0.198	0	2	n/a	n/a	n/a	n/a	n/a	n/a
Midline only	Midline 19	Learners write strokes and curves of letters	0.3	0.807	0	4	n/a	n/a	n/a	n/a	n/a	n/a

*Refers to the significance of the difference in the midline and endline means (independent t-test; independent t-test assuming unequal variance, except where inequality of variance could not be confirmed at the 5 percent significance level).

ANNEX 3. DETAILED EVALUATION METHODS

This annex provides additional detail regarding sample size calculations, achieved sample sizes, design effect, and weights; data collection tools and methods; instrument reliability and validity; inter-rater reliability; and evaluation limitations.

SAMPLE

SAMPLING DESIGN, STAGES, AND STAGE 1 SAMPLING FRAME

TTL's interventions target the *population* of all community schools in its six-province intervention area, all of which should benefit from TTL's interventions. The construction of a control or comparison group was impossible, as there were no community schools with the same languages of instruction that were not already targeted by the intervention. Consequently, the sample drew from the population of registered community schools in these provinces; there is no separation between "intervention" and "comparison" schools in the sample.

[Exhibit 77](#) summarizes the stages of the sampling design, as described in [3.1 Evaluation Design](#).

Exhibit 77: Sampling Stages

Stage of Sampling	Population	Type of Sampling	Sample
Stage 1 (clusters)	TTL intervention schools in each of six provinces (six samples)	Probability proportional to size; Lusaka Province stratified by rural/urban	School
<i>In schools with more than 1 section of grade 2: intermediary stage</i>	Grade 2 classes	Simple random	Class
Stage 2	Learners in selected grade 2 class	Simple random, stratified by sex	Learners

At each evaluation phase, MOGE annual school census data from the previous year provided the stage 1 sampling frame (schools). Thus, the 2011 annual school census was the frame for the baseline (2012), the 2013 census provided the frame for midline (2014), and the 2015 annual school census data provided the endline sampling frame. At endline, after removing schools that lacked enrollment figures, school type (e.g., government, community, private, and grant-aided) or did not conform to the language of instruction for that province (see [Exhibit 12](#) under sampling design in Section 3.1), the frame contained 1,523 schools meeting the sampling criteria. This figure is lower than the project's estimated number of community schools in its catchment areas, because MOGE data only include registered community schools that have completed and returned their annual school census forms during the previous year and due to application of the restrictions applied to the sampling frame, described above.

SAMPLE SIZE PROJECTIONS (POWER CALCULATIONS)

[Exhibit 78](#) shows the target sample sizes for learners and schools at baseline, midline, and endline. These targets were initially calculated prior to baseline for two provinces only, but were revised midway

through data collection to respond to changes to the project implementation schedule that opted for an immediate scale-up instead of the original phased rollout to provinces. As a result, targets were revised upward before midline to account for changes to the evaluation design, as specified in the Midline Evaluation Implementation Plan. The table below presents the updated power analysis conducted prior to midline, and thus presents the target statistical power for comparison of midline to endline data.

The statistical test informing power analysis for learner outcomes (measured by EGRA) is an independent t-test, one tailed, $d=.3$, $\alpha=.05$, which was projected to provide 95.9 percent power to comparisons in sex-aggregated, provincial-level means from midline to endline and 77.3 percent power for differences in sex-disaggregated means across the same period.

Exhibit 78: Target Sample and Power by Evaluation Phase

	Per Province				Six Provinces	
	Schools (Sample Stage 1; Cluster)	Learners (Sample Stage 2)			Total Schools (Classroom Observation)	Total Learners (EGRA)
		Male	Female	Total		
Baseline	10	100	100	200	60	1,200
Midline	17	128	128	256	102	1,536
Endline	17	128	128	256	102	1,536
Power (midline to endline)	n/a	77.3%		95.9%	79.4%	n/a

The target sample for classroom practice outcomes is based on the number of schools sampled at stage 1 (one lesson observed per school). The statistical test is an independent t-test, one tailed, $d=.35$, $\alpha=.05$. The statistical power is only for comparison between midline and endline evaluations at the provincially aggregated level; replacement of the classroom observation tool at midline renders comparison with the baseline inappropriate.

ACHIEVED SAMPLE AND DESIGN EFFECT

The detailed information on achieved sample and intra-class correlation coefficients in this section is included to benefit future sample designs for community school data collection in Zambia.

[Exhibit 79](#) and [Exhibit 80](#) provide the baseline and midline sample overviews, respectively; the information for endline is included in [Section 4](#). At baseline, sample expectations were met only in the two provinces originally scheduled for 2012 data collection (Eastern and Lusaka); targets were missed in the other provinces due to the closure of the school term and the delayed start to data collection, resulting from the changes to the implementation and data collection schedule described above. Commensurate with the evaluation design at the time, learners were only sampled from 73 schools, while head teacher questionnaires were administered at 194 schools; this feature of the evaluation design was modified at midline, leading to alignment of the EGRA and head teacher questionnaire samples at the school level. At midline, sample expectations were met (or surpassed) in all provinces except Muchinga, which achieved a sample of only 16 schools (230 learners), due to premature school closing for caterpillar collection in Mpika District. Data collection was extended in Muchinga Province, but the onset of rains constrained the ability to meet the target sample.

Exhibit 79: Baseline Sample Overview

Province	Schools: Sample Stage 1	Learners Participating in EGRA: Sample Stage 2				Head Teacher Questionnaire Respondents			
		Male	Female	Unknown	Total	Male	Female	Unknown	Total

Central	10	90	84	1	175	23	2	1	26
Copperbelt	10	89	105	1	195	15	12		27
Eastern	18	119	142	10	271	9	38		47
Lusaka	17	103	129	2	234	19	24	1	44
Muchinga	9	60	62		122	22	3		25
Southern	9	68	76		144	18	7		25
Total	73	529	598	14	1,141	135	57	2	194

Exhibit 80: Midline Evaluation Sample Overview

Province	Schools: Sample Stage 1*	Learners Participating in EGRA: Sample Stage 2			Teachers Participating in Classroom Observation				Head Teacher Questionnaire Respondents**		
		Male	Female	Total	Male	Female	Unknown	Total	Male	Female	Total
Central	17 (3)	147	131	278	8	8	1	17	11	6	17
Copperbelt	17 (6)	150	158	308	5	11	1	17	10 (2)	7	17 (2)
Eastern	17 (2)	146	146	292	10	3	4	17	12 (2)	5	17 (2)
Lusaka	18 (2)	141	141	282	4	13	1	18	10 (1)	8 (1)	18 (2)
Muchinga	16 (1)	107	123	230	11	4	1	16	14	2	16
Southern	17 (1)	143	150	293	12	3	2	17	15	2	17
Total	102 (15)	834	849	1,683	50	42	10	102	72 (5)	30 (1)	102 (6)

* Parentheticals denote schools with more than one grade 2 class.

** Parentheticals denote participants with other titles.

The sample identified alternate schools (“reserve units”) for use in the case of a sampling “refusal,” which occurs when data cannot be collected from a selected school for any reasons (e.g., the school has closed, is inaccessible, or does not consent to the procedures). Refusals were removed from the sampling frame after data collection. At endline, 12 of the selected schools (11 percent of the sample) were refusals; reasons for refusal are presented in [Exhibit 81](#) and were mostly related to errors in the school census data.

Exhibit 81: Endline Stage I Sampling Refusals by Province

Province	Number	Reason
Central	4	All four schools’ type was misspecified in the sampling frame; schools were government schools.
Copperbelt	0	
Eastern	0	
Lusaka	5	Two schools were no longer in operation. Three schools’ official language of instruction was ChiTonga and could not be tested in CiNyanja.
Muchinga	1	Two entries in the sampling frame were duplicates referring to the same school.

Southern	2	One school's type was misspecified in the sampling frame; the school was a government school. One school was no longer in operation.
Total	12	

At midline, four sampled schools, two each in Central and Muchinga provinces (4 percent of the sampled schools) were refusals as a result of having been upgraded from community schools to government schools since the 2013 school census data were reported.

The intra-class correlation coefficients provide information on the design effect resulting from the cluster-based sample. These data are provided purely to inform future sampling. [Exhibit 82](#) presents letter sound (task 2) intra-class correlation coefficients and [Exhibit 83](#) presents oral passage reading (task 5a). Although these coefficients are typically only presented for oral passage reading, the floor effect of task 5a, particularly at baseline, may have a distorting impact. Consequently, task 2 provides additional nuance that may prove useful. Coefficients are presented by language, as this is how the EGRA is administered, and for each evaluation phase, as coefficients' evolution over time can often prove unpredictable.

Exhibit 82: Letter Sounds (task 2) Intraclass Correlation Coefficients by language

Language	Baseline	Midline	Endline	All Phases
ChiTonga	0.04157	0.03064	0.54683	0.52447
CiNyanja	0.30015	0.26086	0.39731	0.31898
iCiBemba	0.25856	0.23602	0.27531	0.30953
All Languages	0.30176	0.28800	0.37518	0.35646

Exhibit 83: Oral Passage Reading (task 5a) Intraclass Correlation Coefficients by language

Language	Baseline	Midline	Endline	All Phases
ChiTonga	0.1423	0.08155	0.60006	0.50257
CiNyanja	0.41593	0.16386	0.29478	0.23763
iCiBemba	0.16035	0.13677	0.21754	0.20576
All Languages	0.26706	0.15352	0.2848	0.2417

WEIGHTING

All EGRA results presented in this report are weighted using probability weights, except where otherwise noted.

The baseline stage 1 sample was selected using simple random sampling. At midline and endline, a subset of the baseline sample schools was retained in the sample (selected with probability equal to 1) to ensure a stable subsample of schools with data across all three phases. The high turnover among community schools, with new schools opening and others closing each year, combined with the gaps in the MOGE's annual school census data, means the sampling frame was unstable across the three evaluation phases. Consequently, this subset of schools with data for all three phases served two purposes:

- It provides a longitudinal group of schools that can be further analyzed in future studies to understand the long-term dynamics of community schools.
- It ensures a group of schools that benefited from the full course of the intervention.

In order to be retained in the midline and endline samples, schools in this subsample had to appear in the respective sampling frame for that evaluation phase and meet the same criteria as other schools in the frame; in other words, the school still had to be a valid sampling unit at the later phase. This subset of schools is hereafter referred to as the “certainty sample.” In reality, the schools were selected randomly at baseline and are only certainty sampling units for purposes of calculating probability weights.

Final sampling weights were calculated as the inverse of the ultimate probability of selection for each learner, taking into account all sampling stages, including the intermediary stage and the certainty sampling, where applicable. The probability of selection of a learner is calculated as follows:

$$p(\text{learner}) = n_h \frac{M_{hi}}{M_h} \frac{m_{hi}}{M'_{hi}}$$

Where:

n_h = Number of schools to be chosen per province, excluding schools in the certainty sample at endline¹⁹

M_{hi} = Total number of learners in grade 2 in the sample school as shown in the frame

M_h = Total number of learners in grade 2 in the province (excluding those in the certainty sample at endline)

m_{hi} = Total number of grade 2 learners selected in the class (in classes with less than 10 learners per sex, all learners of that sex were selected)

M'_{hi} = Total number of grade 2 learners observed in the sample school prior to the interview. Due to the time difference between reporting of annual school census data informing the sampling frame and the data collection for this evaluation, in most cases this number does not coincide with M_{hi} .

The initial weighting factor is the inverse of the probability of selection, and for this sampling design, is also the final weighting factor:

$$W(\text{learner}) = \frac{1}{n_h} \frac{M_h}{M_{hi}} \frac{M'_{hi}}{m_{hi}}$$

¹⁹ In calculating the midline probability weights, it was decided not to give these schools a certainty probability, due to imperfect records in the MOGE school census data (i.e., unrecorded closed schools, missing enrollment data, and the large number of certainty schools from baseline that could not be identified in the midline frame). In these cases, it is recommended that all the schools in the sample represent all the schools in the frame in order to spread the sample representation more evenly. Thus, at midline, schools in the certainty sample were added to the rest of the pool for weight calculation and were not chosen with probability 1. Treating these schools as if they had been chosen the same as the rest of the schools in this way is reflected in the formula used to determine the probability of selection of the school and, subsequently, its inverse, the weight.

The formula reflects the two-stage sample design described at the beginning of this annex and the certainty subsample where applicable.

TOOLS AND METHODS

EARLY GRADE READING ASSESSMENT

The EGRA is a literacy assessment focusing on foundational pre-reading and reading competencies proven through research and instrument validation to be necessary for early reading and literacy acquisition. The EGRA is administered orally to one learner at a time in the language of literacy instruction.

TTL shares a similar EGRA tool with the USAID Read to Succeed Project (working with government schools), and the ECZ’s 2014 grade 2 National Assessment. Harmonization of the EGRA tool across these three entities facilitates comparison of data across the three EGRAs, although this comparison is beyond the scope of the TTL endline evaluation, due to each organization’s data preparation schedule.

The EGRA used in 2014 and 2016 follows the general structure of the 2012 baseline EGRA, with two key structural modifications. The following modifications were made for the midline to respond to MOGE and TTL project priorities and the new MOGE-issued performance level descriptors, which provided concrete benchmarks for early grade literacy performance:

- **Inclusion of orientation to print in all provinces:** Orientation to print was introduced during the second round of baseline data collection in Central, Copperbelt, Muchinga, and Southern provinces in 2012, but had not been done in the first round of baseline data collection in Eastern and Lusaka provinces in 2012.
- **Substitution of English vocabulary with English listening comprehension** in order to more accurately capture learner understanding of the English language (the official language in Zambia).

Additionally, each phase of the TTL evaluation (baseline 2012, midline 2014, and endline 2016) used slightly modified EGRA reading passages and associated comprehension questions to measure grade 2 learners’ oral reading fluency (task 5a) and reading comprehension (task 5b). Modification of passages ensures that differences in learners’ reading ability detected over time were not due to the readers’ familiarity with the instrument.²⁰ Equating procedures adjust for potential differences in difficulty between the versions of the instruments used at each phase so that the scores can be compared across these phases. [Exhibit 84](#) presents the equating methods and score conversions for oral reading fluency and reading comprehension in each of the three TTL languages to convert endline scores to the baseline scale. More details on the equating process for the endline instruments can be found in [Annex 4](#). The equating ratios used to convert the midline scores to the baseline scale are provided in the 2014 Grade 2 National Assessment (RTI International 2015).

Exhibit 84: 2016 TTL to 2012 Oral Reading Fluency and Reading Comprehension Equating Methods and Score Conversion

Language	CiNyanja	ChiTonga	iCiBemba
----------	----------	----------	----------

²⁰ For example, the baseline passages served as the basis for TTL teacher materials used in classrooms.

	Oral reading fluency (task 5a)		
Method	Identify	Mean	Mean
Equating Score	1.000	0.898	1.093
	Reading Comprehension (task 5b)		
Method	Circle-arc	Circle-arc	Circle-arc
Number Correct: 0	0.00	0.00	0.00
1	0.41	0.77	0.77
2	1.15	1.66	1.66
3	2.15	2.66	2.66
4	3.41	3.77	3.78
5	5.00	5.00	5.00

Through use of these score conversions on the 2016 EGRA data and the score conversion provided in the 2014 Grade 2 National Assessment Report on the 2014 data, EGRA results are comparable across all three evaluation phases for the following four tasks:

- Task 2: Letter sound knowledge
- Task 3: Non-word decoding
- Task 5a: Oral passage reading
- Task 5b: Reading comprehension.

The midline and endline EGRA data are additionally comparable for the other three tasks.

COMMUNITY SCHOOL HEAD TEACHER QUESTIONNAIRE

In each sampled school, the evaluation team was to interview the head teacher using this questionnaire. If the head teacher was absent on the day of the school visit, another school representative was interviewed instead. The evaluation team made minor updates to the baseline version of the community school head teacher questionnaire to account for changes in project implementation, enable measurement of exposure to TTL interventions, and improve question clarity and ease of administration for administrators and respondents. The evaluation team retained core questions at both midline and endline, for comparison to baseline data to respond to evaluation question 2 and to help explain learner EGRA scores. This questionnaire was modeled on Snapshot of School Management Effectiveness tools and captured key data such as school enrollment, external support types and sources, and parental engagement.

CLASSROOM OBSERVATION PROTOCOL

At baseline, the evaluation team collected data on classroom literacy practice using an adapted version of Education Development Center, Inc.'s Standards-based Classroom Observation Protocol for literacy, known as SCOPE. This tool was replaced after the baseline for the following reasons:

- Baseline data collectors found it difficult to administer.
- The tool focuses on higher-order pedagogical practices, not the specific early grade literacy instructional methods promoted by the TTL training program.
- The baseline data exhibited very low inter-rater reliability scores and weak correlation with EGRA scores.

Consequently, the evaluation team developed the TTL-specific classroom observation protocol for the midline evaluation to capture literacy instruction practices. While this change of tools limits comparison of teacher practice data at midline and endline to the baseline values, that calculation would be of limited utility to the evaluation, as the baseline tool was not appropriate for measuring change related to TTL interventions.

The benefit of this replacement is that TTL technical specialists, USAID, the MOGE, and future project implementers gain more specific literacy insights into current classroom instructional practice and uptake of TTL methods by community school teachers. At midline and endline, the tool proved capable of generating targeted descriptive data on instructional delivery with high reliability. In addition, the comparison possible between midline and endline reveals changes in the classroom over the last 2 years of the project.

The classroom observation protocol includes a post-observation teacher questionnaire that provides insight into teachers' beliefs and background. In cases where the observed teacher was also the respondent for community school head teacher questionnaire, the post-observation teacher questionnaire component was omitted to avoid placing an undue time burden upon a single respondent. The evaluation team made minor updates to the midline version of the observed teacher questionnaire to account for changes in project implementation, enable measurement of exposure to TTL interventions, and improve question clarity and ease of administration for administrators and respondents. The evaluation team retained core questions for comparison to midline data to respond to evaluation question 3 and to help explain learner EGRA scores.

While all observation data were disaggregated by sex during analysis to examine differences in uptake between male and female teachers, the sample is not guaranteed to be large enough that these disaggregations are sufficiently powered to conduct statistical comparison between subgroups.

TOOL RELIABILITY AND VALIDITY

INTER-RATER RELIABILITY

The classroom observation protocol and EGRA are quantitative measures based on subjective assessments by data collectors. Inter-rater reliability testing assesses the level of consistency between data collectors; it is, therefore, a key component of ensuring data quality for such tools. The evaluation team assessed inter-rater reliability during data collector trainings for both EGRA and the classroom observation protocol for all trainees, and additionally assessed inter-rater reliability on a random sub-sample of classroom observations during data collection in the field. Three measures of reliability and agreement are used: raw percentage agreement (only used in training scenarios), the intraclass correlation coefficient, and the Kappa statistic.

During trainings, raw percentage agreement measured each data collector's agreement against a single master score (a "gold standard"). Tool developers and training facilitators carefully determined this "gold standard" score. Training facilitators used the raw percentage agreement during training for formative purposes and at the end of training for selecting quality assessors. At midline and endline trainings, EGRA assessor agreement scores on the final day of training ranged from 78–97 percent agreement, with a 90 percent threshold to conduct EGRA in the field. Only one trainee failed to meet this threshold. For the classroom observation protocol, trainees were required to obtain an 80 percent agreement threshold to be selected as a data collector. Scores range from 79–93 percent on the final day of training, with only one trainee failing to meet the threshold. Trainees who did not meet the thresholds were not part of the data collection teams. [Exhibit 85](#) and [Exhibit 86](#) provide additional reliability statistics for the last day of training for the final group of trainees selected to serve as data collectors. Intraclass correlation coefficient values above 0.75 are generally considered to indicate

excellent reliability; thus, these results indicate high consistency of scores among trainees (RTI International 2015: 206).

Exhibit 85: Inter-rater Reliability during EGRA Training

Language	Intraclass Correlation Coefficient	Kappa Statistic
ChiTonga	0.997	0.389
CiNyanja	0.999	0.524
iCiBemba	0.933	0.543

Exhibit 86: Inter-rater Reliability for Classroom Observation Protocol Training and Field Sample

	Intraclass Correlation Coefficient	Kappa Statistic
Training Assessment	0.970	0.346
Field Sample	0.878	0.648

Classroom observation inter-rater reliability was also tested throughout field data collection at both midline and endline to ensure continued consistency in tool application (guarding against “assessor drift”) and that the actual data would maintain high reliability. At endline, 16 percent (n=17) of classroom observations were sampled for inter-rater reliability testing during data collection. For each of these 17 lessons, the data collection team lead (who participated in the full training on all data collection tools and achieved the same reliability thresholds during training) independently and simultaneously observed the same lesson as that team’s primary classroom observation protocol administrator. Each pair of scores was then compared to determine inter-rater reliability. [Exhibit 86](#) above presents the inter-rater reliability for these 17 pairs using two different methods: the intraclass coefficient (two-way mixed model specifying absolute agreement) and the Kappa statistic (Cohen’s method). The results indicate excellent inter-rater reliability. Results from parallel procedures conducted at midline can be found in the Midline Evaluation Report (Annex 2, pages 61–63). The results from both phases confirm that inter-rater reliability remained high throughout data collection and thus provides evidence supporting the consistency of the data used in this report.

INSTRUMENT INTERNAL RELIABILITY AND VALIDITY

EGRA: Cronbach’s alpha values, presented in [Exhibit 87](#), indicate that the EGRA instruments exhibit strong internal reliability for all three language versions and at all phases. All tools exhibit “acceptable” levels of internal reliability (0.70 and higher) at baseline and midline, and increasing a “very good” reliability (greater than 0.80) level at endline (RTI International 2015). The internal reliability of the instrument has improved over time, which may reflect continued work to improve the Zambia EGRA instruments and reduced administrator error over time. TTL’s Cronbach’s alpha values align with instrument reliability findings from the Grade 2 National Assessment conducted in 2014, which also assessed instrument validity and found positive results (RTI International 2015).

Exhibit 87: EGRA Internal Reliability Check (Cronbach’s alpha)

Language	Baseline	Midline	Endline	All Phases
CiNyanja	0.7074	0.7272	0.8107	0.8012
ChiTonga	0.733	0.7297	0.8477	0.8298

iCiBemba	0.8175	0.8046	0.8135	0.824
----------	--------	--------	--------	-------

Classroom Observation Protocol: Internal reliability of the classroom observation protocol was assessed by calculating Cronbach's alpha on the midline data subsequent to data collection. These results are presented in the Midline Evaluation in [Exhibit 75](#) in [Annex 2](#) for the sub-scales formed by each of the domains. While some sub-scales demonstrated acceptable reliability, the number of items in each domain is small and, overall, Cronbach's alpha values were observed to be low. This likely reflects that the underlying constructs the tool measures are complex and multidimensional; indeed, TTL literacy experts and the evaluation team intentionally designed the tool to capture distinct aspects of literacy lessons that build different reading competencies. These results support the presentation in this report of classroom observation data by item, rather than by domain or for the overall scale.

LIMITATIONS

DESIGN LIMITATIONS

Independently pooled cross-sectional designs depend on data from more than one period of time. Throughout this report, findings reporting change or comparing 2016 to 2014 and to 2012 draw on baseline datasets. Cross-sectional designs are adept at capturing changes in indicators over time, but cannot statistically attribute changes to a specified intervention. Through the broader 5-year evaluation mixed-methods evaluation approach, TTL evaluations seek to build evidence for attribution, but this should not be conflated with statistical attribution.

Baseline sample sizes were constrained by USAID changes to the intervention timeline, as described above. In addition, sampling frames were limited by incomplete information regarding the community schools' population in the census data.

Linguistic diversity in Zambia means reading score population estimates at provincial level do not include language minority groups in Central, Lusaka, and Muchinga provinces.

TOOL LIMITATIONS

EGRA: Insufficient baseline data exist for task 4 (orientation to print) and for task 6 (English language listening comprehension) in all provinces. Task 1 language of instruction listening comprehension passages were altered during the 2014 Grade 2 National Assessment process, but no equating was conducted, so the results cannot be accurately compared to 2012.

EGRA passages are language-specific and the languages themselves differ in difficulty, affecting the speed of reading acquisition. Thus, EGRA results cannot be compared across the three languages of instruction in which TTL works. The MOGE (2015a) benchmarks for grade 2 early grade literacy, however, are the same for all languages, irrespective of these differences.

EGRA passages are designed to be grade-appropriate, but because the MOGE benchmarks lack a common definition of grade-appropriate text, there is room for differences in interpretation in the meaning of "grade level." Furthermore, the 2016 EGRA has been equated with the 2012 EGRA, which predated the benchmarks. This increases the possibility for differences in defining what is truly grade-appropriate. Because the standards are the same for all seven languages of instruction but languages inherently differ in difficulty, there is also the risk that the standards may be inappropriately high for more challenging languages.

Classroom Observation Protocol: The replacement of the baseline classroom observation tool removed the ability to compare teacher practice between baseline and midline. However, the new tool

corresponds more closely with TTL-promoted classroom instructional practice and EGRA test items and offers increased explanatory power in understanding variation in learner reading outcomes. At endline, additional changes to some of the classroom observation criteria mean that not all criteria are available to compare between midline and endline.

DATA LIMITATIONS

TTL's commitment to engaging MOGE officials and TTL staff as data collectors built their capacity over time, but using data collectors who did not have substantial prior experience collecting classroom observation and EGRA data affected data quality at baseline. As detailed above, inter-rater reliability results show that this challenge was effectively mitigated in the endline. However, the endline evaluation is subject to the data quality constraints of earlier evaluation phases in comparisons across time.

Classroom Observation Protocol: In 28 cases (27 percent of the sample), the teacher observed was the same as the community school head teacher questionnaire respondent. For these cases, the post-observation teacher questionnaire was omitted. Consequently, teacher self-reflections on teaching practice and attitude data are missing for these 28 cases.

Community School Head Teacher Questionnaire: Recall issues are particularly likely for questions related to training: it is possible that head teachers did not remember every training session teachers at the school attended, which might result in underreporting. Schools may have a “vested interest” in underreporting MOGE support received, with the hope of getting more support. Responses related to grants were attributed to the MOGE if they were from any government agency, including Constituency Development Fund grants. The reason for this is that schools might not know or recall the actual government agency providing the grant, and MOGE district offices often support schools in getting Constituency Development Fund grants, which originate through the office of the local Parliamentarian.

MOGE Self-Administered Survey Questionnaire: Because this questionnaire was self-administered, there is a risk that responses were based on recall or that requested data was missing or undocumented and could not be accurately reported. Additionally, the MOGE may have a “vested interest” in reporting high levels of support.

As noted in [Section 3.2](#), the achieved response rate for this tool provides information for 93 percent of the 105 schools sampled in the endline. It is possible that divergence between the MOGE- and head teacher-reported levels of MOGE support to community schools is partially a result of this response rate.

ANNEX 4. EQUATING OF EGRA PASSAGES

SUMMARY

This annex presents information from a study on instrument equivalency for the purpose of equating results based on modified reading passages used for assessing two EGRA tasks: task 5a (oral reading fluency) and task 5b (reading comprehension). Results are presented for each of the three Zambian languages of instruction for which the endline presents EGRA data: ChiTonga, CiNyanja, and iCiBemba. Each phase of the TTL evaluation (baseline 2012, midline 2014, and endline 2016) used slightly modified EGRA reading passages to mitigate risks of passage leakage and ensure that differences in learner reading ability detected over time were not due to familiarity with the instrument. Because the TTL endline uses EGRA data to calculate change in reading ability at the population level across time, equating procedures adjust for potential instrument differences so the scores can be compared across these phases. This study used a single-group counterbalanced design, in which each learner is his or her own matched pair, for data collection and equating analysis.

The baseline and midline passages were developed in 2012 and 2014. The TTL endline passages were developed in June 2016 at a workshop facilitated by the TTL evaluation team, in conjunction with language and assessment experts from the MOGE Curriculum Development Center and the ECZ. TTL and the USAID Read to Succeed project share EGRA instruments to ensure comparability of results across projects, so Read to Succeed participated in this workshop to simultaneously facilitate adaptation of EGRA instruments in the languages for which it conducts assessment, ensuring the use of parallel processes across the two projects.

This annex also presents equating findings for the new reading passages the ECZ developed in 2016, which will be used in the next Grade 2 National Assessment. The ECZ passages were finalized at the same workshop and incorporated into the pilot data collection described below, along with the TTL passages. The ECZ needed these passages equated to facilitate comparison of the next Grade 2 National Assessment with the 2014 Assessment, which used the same reading passages as the TTL midline. Because the TTL endline assesses differences from the baseline, the 2016 TTL version is converted to the 2012 scale. The 2016 ECZ version is converted to the 2014 passages, because this is the point of reference for the national assessments. The two reading passage versions created in 2016 are referred to here as “2016 TTL” and “2016 ECZ.”

The TTL evaluation team tested five equating methods (identity, mean, linear, equipercentile, and circle-arc) to select the most robust method. The final equating methods selected are listed in [Exhibit 88](#). [Exhibit 89](#)

[Exhibit 89](#) and [Exhibit 90](#) provide the equating ratios and score conversions necessary for comparing data between various test versions (and evaluation phases); applying the transformations specified equates the corresponding 2016 data to the baseline score. Oral reading fluency is converted using equating ratios for consistency with the midline equating methodology ([Exhibit 89](#)); reading comprehension is converted using a fixed score conversion. [Exhibit 90](#) presents the equating recommendations for the TTL endline passages in comparison to learner scores at baseline. [Exhibit 91](#) presents equating recommendations for the ECZ passages.

Exhibit 88: Equating Method for EGRA Instrument by Language

	Oral Reading Fluency		Reading Comprehension	
	2016 TTL → 2012	2016 ECZ → 2014	2016 TTL → 2012	2016 ECZ → 2014
ChiTonga	Mean	Identity	Circle-Arc	Circle-Arc
CiNyanja	Identity	Identity	Circle-Arc	Circle-Arc
iCiBemba	Mean	Identity	Circle-Arc	Identity

Exhibit 89: Oral Reading Fluency Equating Ratios by EGRA Language Instrument

	Oral Reading Fluency	
	2016 TTL → 2012	2016 ECZ → 2014
ChiTonga	0.898	1.000
CiNyanja	1.000	1.000
iCiBemba	1.093	1.000

Note: 1.000 indicates identity equating.

Exhibit 90: Reading Comprehension Score Conversions, 2016 TTL to 2012 Instruments

2016 TTL to 2012 Reading Comprehension Score Conversions			
	Reading Comprehension		
	ChiTonga	CiNyanja	iCiBemba
0	0.00	0.00	0.00
1	0.77	0.41	0.77
2	1.66	1.15	1.66
3	2.66	2.15	2.66
4	3.77	3.41	3.78
5	5.00	5.00	5.00

Exhibit 9I: 2016 ECZ to 2014 Reading Comprehension Score Conversions

2016 ECZ to 2014 Reading Comprehension Score Conversions			
	Reading Comprehension		
	ChiTonga	CiNyanja	iCiBemba*
0	0.00	0.00	0
1	1.13	0.88	1
2	2.19	1.82	2
3	3.19	2.82	3
4	4.13	3.88	4
5	5.00	5.00	5

* Indicates identity equating.

EQUATING EXPLAINED

Due to security or instrument validation reasons, it is common for programs that administer standardized tests to administer parallel forms of a test at different times. Even if test developers construct test forms whose content and statistical characteristics are the same, the forms can have different difficulty levels. Equating is a statistical procedure used to convert scores from multiple forms or variations of a test to the same common measurement scale (Holland and Dorans 2006; Kolen and Brennan 2004). Comparable test scores are an important consideration for consistent tracking of results over time, as in the case of EGRA scores in measuring results for USAID's early grade reading programs.

Equating methods can be classified under the two broad theories of classical test theory and item-response theory. In this equating exercise, methods based on classical test theory have been used due to the limitations of the small sample and the nature of the data. The five classical test theory methods are:

- **Identity equating:** A score on one test form is treated as an exact match to the same score on another test form; thus, no conversion is applied prior to comparing scores. Identity equating is the same as not equating at all; however, equivalency of test forms is verified prior to using this approach through equating analysis.
- **Mean equating:** The difficulty of one test form is differentiated from another by a set amount, which is the mean difference between the two forms.
- **Linear equating:** Scores are considered equivalent when the scores on two test forms have the same standard-score deviations (Angoff 1984).
- **Equipercentile equating:** Two scores on two forms may be considered equivalent if their corresponding percentile ranks in any given group is equal (Kolen 1988; Kolen and Brennan 2004; Livingston 2004).
- **Circle-arc equating:** This model does not assume a linear relationship between test forms (Livingston and Kim 2009b, 2010). It determines a *circle-arc* by three points, the lower endpoint as the lowest meaningful test score in each form, the upper endpoint as the maximum possible score for the test form, and the middle point on the curve is the point in the middle of the distribution of scores. If these three points are on a straight line, this line is called the estimated equating curve. The sum of the linear and curvilinear components composes the circle-arc equating function.

As described under [Analysis Procedure](#), the standard error of equating is used as a criterion for comparing results among these five different equating methods. The standard error of equating is “the standard deviation of equated scores over replications of an equating procedure in samples of examinees” (Kolen and Brennan 2004: 211). As such, the standard error of equating is measuring the random error introduced by the equating and serves as an index for determining the accuracy of the equating function. The most appropriate equating method in a particular context is determined by identifying the one with the lowest standard error of equating.

2016 EGRA EQUATING DATA COLLECTION

The evaluation team collected the data used to conduct the equating analysis described in this annex from June 20–July 1, 2016, immediately following the instrument adaptation workshop. Consequently, the instrument adaptation was conducted as pre-equating, prior to the full survey implementation that collected the data used in the rest of this endline report.

EGRA assessors participating in the equating exercise were selected from a pool of individuals with deep prior experience conducting EGRA, with TTL and serving as EGRA trainers at the district level. TTL, with support from ECZ and Read to Succeed, conducted a 3-day refresher EGRA training to acquaint EGRA assessors with the pilot versions of the instruments, familiarize data collectors with special considerations of the equating pilot and its data collection procedures, and provide opportunity to practice using the pilot instrument versions. To ensure data collector consistency in using the new version of the instruments, EnCompass conducted two rounds of Inter-rater Reliability Testing. All assessors cleared a 90 percent agreement threshold against a gold standard (answer key) using the raw percentage agreement method for oral reading fluency, and achieved 100 percent agreement on reading comprehension.

To ensure an adequate distribution of scores representative to the population and to increase the usable sample size (non-zeros), the equating exercise assessed learners across grades 2–4 from large government schools known to have good early grade reading performance. Details of the final pilot sample are shown in [Exhibit 92](#). All learners participating in this equating exercise read all four passages: baseline (2012), midline (2014), and two 2016 versions (TTL and ECZ), given in a randomized order. Learners then answered up to five reading comprehension questions, based on how far through the passage they had read; this reflects the standard EGRA administration procedure used globally and in Zambia (including the TTL baseline, midline, and endline and the 2014 Grade 2 National Assessment).

Exhibit 92: Number of Learners Sampled in Pilot Assessment, by Language and Grade

	Language and Grade								
	ChiTonga			CiNyanja			iCiBemba		
	G2	G3	G4	G2	G3	G4	G2	G3	G4
Learners sampled	157	155	111	193	118	59	246	91	30
Total (including zeroes)	423			370			367		
Learners scoring zero or with data missing on any of the four passages	33	4	11	35	11	5	37	17	4
Learners with fewer than 5 seconds elapsed	0	0	0	0	0	0	1	1	0

ANALYSIS PROCEDURE

The analysis and recommendations follow two guidelines. First, if the mean difference between tasks on the two forms is less than 0.1 of the raw score standard deviation unit, identity equating is recommended (Hanson, Zeng, and Colton 1994; Skaggs 2005; RTI International 2015). Second, the standard error of equating for the selected method should also be less than 0.1 of the raw score standard deviation (Kolen and Brennan 2004). The final equating procedure selected for each task (as seen in [Exhibit 88](#)) is the method resulting in the lowest standard error of equating.

ORAL READING FLUENCY

Oral reading fluency is measured by correct words per minute. There is no fixed maximum score for correct words per minute, since learners can finish the passage in less than a minute. With four forms per task (2012, 2014, 2016 ECZ, and 2016 TTL), each learner has eight scores. Cases with fewer than 5 seconds elapsed for oral reading fluency were coded as invalid due to assessor error and excluded from the analysis, because the resulting outliers can severely influence mean scores for small samples such as this. For the pairs of oral reading fluency scores, only non-zero scores were included in the equating. High proportions of zero scores violate the normality assumption and bias the equating results.

Based on these guidelines, the equating method selected for oral reading fluency is the mean or identity method, depending on the language ([Exhibit 88](#)). In addition to meeting the two guidelines above, the mean and identity methods are recommended for two primary reasons. First, the reading fluency scores between the two forms exhibit a high linear correlation. Second, the circle-arc method is inappropriate, since there is no fixed maximum oral reading fluency score.

The mean method generates the equating ratios provided in [Exhibit 89](#), which are used to convert endline scores to the baseline scale by multiplying the raw endline score to the appropriate ratio. A common alternative to the equating ratio conversion is the mean difference conversion. This alternative conversion is provided in [Exhibit 93](#) for informational purposes only, but this conversion method is not used in the TTL evaluations, because previous Zambia EGRA results have exhibited a high proportion of zero scores; consequently, zero score analysis has been a central indicator by which to assess progress. **In this context, the evaluation team recommends the mean ratio score conversion, as the mean difference conversion inflates zero scores.**

Exhibit 93: Mean Difference Constant for Score Conversion (Alternative to Equating Ratio Conversion), by Language²¹

	Oral Reading Fluency	
	2016 TTL → 2012	2016 ECZ → 2014
ChiTonga	2.38	0.00
CiNyanja	0.00	0.00

²¹ Using this method, a score is converted to the baseline scale by subtracting the constant from the raw score on the endline form. The constant is calculated as the baseline mean subtracted from the endline mean.

iCiBemba	-1.88	0.00
----------	-------	------

Note: Zero indicates identity equating.

READING COMPREHENSION

The reading comprehension score is the total of five dichotomously coded items (five questions scored as either “correct” or “incorrect”), based on the oral reading passage. Thus, the reading comprehension score ranges from 0–5. The recommended equating method for the reading comprehension variable is the circle-arc or identity method, depending on the language ([Exhibit 88](#)). In addition to the guidelines above, this method is recommended because the data are not purely linear and the variable has fixed minimum and maximum values. The circle-arc method generates a raw-to-scale conversion table that list each raw score point on the endline form and its equivalent score on the baseline scale (see [Exhibit 90](#) and [Exhibit 91](#)).

SUMMARY STATISTICS BY LANGUAGE

[Exhibit 94](#) and [Exhibit 95](#) show the summary statistics for the oral reading fluency and reading comprehension tasks, by language and passage. Consistent with the first guideline above, raw score mean difference against the standard error (here defined as 10 percent of the standard deviation) was calculated to determine whether equating was necessary.

Exhibit 94: Mean, Standard Error, and Sample Size for 2016 TTL and 2012 Passages by Language

Language		Oral Reading Fluency		Reading Comprehension	
		Endline Passages 2016 TTL	Baseline Passages 2012	Endline Passages 2016 TTL	Baseline Passages 2012
ChiTonga	Number of students	423 (28)	423 (40)	395 (94)	384 (104)
	Mean	20.98	18.71	1.44	1.19
	Standard error	1.19	1.17	0.10	0.11
CiNyanja	Number of students sampled	370 (42)	370 (42)	331 (108)	328 (185)
	Mean	16.27	15.64	1.32	0.63
	Standard error*	1.18	1.11	0.12	0.09
iCiBemba	Number of students	361 (45)	361 (47)	318 (69)	316 (98)
	Mean	17.55	19.13	1.47	1.19
	Standard error	1.19	1.23	0.11	0.10

Note: Numbers in the parentheses represent for the numbers of learners with zero scores.

** Here defined as 10 percent of raw score standard deviation.*

Exhibit 95: Mean, Standard Error, and Sample Size for 2016 ECZ and 2014 Passages, by Language

		Oral Reading Fluency	Reading Comprehension
--	--	----------------------	-----------------------

Language		Endline Passages 2016 ECZ	Baseline Passages 2014	Endline Passages 2016 ECZ	Baseline Passages 2014
ChiTonga	Number of students	422 (34)	423 (31)	389 (129)	392 (114)
	Mean	20.58	20.17	1.33	1.48
	Standard error	1.20	1.26	0.13	0.13
CiNyanja	Number of students	370 (22)	370 (28)	346 (62)	342 (68)
	Mean	20.08	19.66	1.95	1.80
	Standard error	1.34	1.34	0.15	0.14
iCiBemba	Number of students	361 (37)	361 (36)	326 (128)	326 (119)
	Mean	22.96	23.52	1.30	1.41
	Standard error	1.38	1.41	0.14	0.14

Note: The numbers in the parentheses represent for the numbers of learners with zero scores.

DETAILED RESULTS BY LANGUAGE: OVERVIEW

The following sections provide detailed descriptive statistics and methodological details for each of the three languages tested. Below is the generic “table of contents” for each of these sections.

LIST OF CONTENTS FOR REPORT BY LANGUAGE

Part I. Summary of Demographic Variables

Part II. Summary Statistics for the Two Reading Tasks

Part III. Equate Results for the Two Reading Tasks

Part V. Summary Results

Table I.1: Frequency for Sex

Table II.1: Summary Statistics for the Oral Reading Fluency Task

Table II.2: Summary Statistics for the Reading Comprehension Task

Table II.3: Correlation Coefficients of the Oral Reading Fluency and Reading Comprehension

Table III.1: Mean Difference and Standard Error for the Oral Reading Fluency Task

Table III.2: Mean Difference and Standard Error for the Reading Comprehension Task

Table III.3: Equating Result for the Oral Reading Fluency 2016 TTL vs. 2012

Table III.4: Equating Result for the Reading Comprehension 2016 TTL vs. 2012

Table III.5: Equating Result for the Reading Comprehension 2016 ECZ vs.2014

Table V.1: Mean Ratios for the Oral Reading Fluency Task

Table V.2: Mean Ratios for the Reading Comprehension Task

Figure I.1: Histogram for Age

Figure II.1: Distributions for Oral Reading Fluency Task

Figure II.2: Distributions for Reading Comprehension Task

Figure II.3: Scatter Plot of Oral Reading Fluency 2016 TTL vs. 2012 and 2016 ECZ vs. 2014

Figure III.1: Cumulative Distribution Plot for the Oral Reading Fluency 2016TTL vs. 2012

Figure III.2: Standard Error of Equating for Oral Reading Fluency 2016 TTL vs. 2012

Figure III.3: Standard Error of Equating for Reading Comprehension 2016 TTL vs. 2012 and 2016 ECZ vs. 2014

DETAILED RESULTS BY LANGUAGE: CINYANJA

PART I. SUMMARY OF DEMOGRAPHIC VARIABLES

Table I.1: Frequency for Sex – CiNyanja

Sex	Female	Male	Total
N	187	138	325

PART II. SUMMARY STATISTICS FOR THE TWO READING TASKS

Table II.1: Summary Statistics for the Oral Reading Fluency Task – CiNyanja

	Fluency 2012	Fluency 2014	Fluency 2016 ECZ	Fluency 2016 TTL
N Valid *	370	370	370	370
Missing	0	0	0	0
Mean	15.64	19.66	20.08	16.27
Standard Deviation	11.091	13.380	13.368	11.769
Minimum	0	0	0	0
Maximum	48	62	60	52

* Indicates the total observations with seconds elapsed greater than 5.

Table II.2: Summary Statistics for the Reading Comprehension Task

	Comprehension 2012	Comprehension 2014	Comprehension 2016 ECZ	Comprehension 2016 TTL
N Valid *	328	342	346	331
Missing	42	28	24	39
Mean	0.63	1.80	1.95	1.32
Standard Deviation	0.85	1.439	1.473	1.238
Minimum	0	0	0	0
Maximum	3	5	5	5

* Indicates the total observations with seconds elapsed greater than 5.

Table II.3: Correlation Coefficients for Two Reading Tasks

	Oral Reading Fluency		Reading Comprehension	
	2016 TTL	2016 ECZ	2016 TTL	2016 ECZ

	vs. 2012	vs. 2014	vs. 2012	vs. 2014
Correlation Coefficient	0.951* (n=370)	0.951* (n=370)	0.634* (n=322)	0.807* (n=339)

* Correlation is significant at the 0.01 level (2-tailed). N is counted as non-missing scores for both tests.

PART III. EQUATE RESULTS FOR THE TWO READING SUBTASKS

Table III.1: Mean Difference and Standard Error for the Oral Reading Fluency Subtask

	2012	2014	2016 ECZ	2016 TTL	Mean Difference	
					2016 TTL vs. 2012	2016 ECZ vs. 2014
Mean	15.64	19.66	20.08	16.27	0.63	0.42
Standard Deviation	11.091	13.38	13.368	11.769		
Standard Error	1.1091	1.338	1.3368	1.1769		

Table III.2: Mean Difference and Standard Error of the Reading Comprehension Subtask

	2012	2014	2016 ECZ	2016 TTL	Mean Difference	
					2016 TTL vs. 2012	2016 ECZ vs. 2014
Mean	0.63	1.8	1.95	1.32	0.69	0.15
Standard Deviation	0.85	1.439	1.473	1.238		
Standard Error	0.085	0.1439	0.1473	0.1238		

Table III.4: Equating Result for Reading Comprehension: 2016 TTL vs. 2012

	Identity			Mean			Linear			Equipercetile			Circle-arc		
	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*
Mean	1.354	0.643	1.354	1.354	0.643	0.643	1.354	0.643	0.643	1.354	0.643	0.634	1.354	0.643	0.865
SD	1.235	0.854	1.235	1.235	0.854	1.235	1.235	0.854	0.854	1.235	0.854	0.865	1.235	0.854	1.021
Skew	0.702	1.230	0.702	0.702	1.230	0.702	0.702	1.230	0.702	0.702	1.230	1.054	0.702	1.230	1.725
Kurt	2.880	3.711	2.880	2.880	3.711	2.880	2.880	3.711	2.880	2.880	3.711	3.633	2.880	3.711	6.547
Min	0.000	0.000	0.000	0.000	0.000	-0.711	0.000	0.000	-0.293	0.000	0.000	-0.223	0.000	0.000	0.000
Max	5.000	3.000	5.000	5.000	3.000	4.289	5.000	3.000	3.163	5.000	3.000	3.312	5.000	3.000	5.000
N	322	322	322	322	322	322	322	322	322	322	322	322	322	322	322
Standard Error of Equating	0.000			0.0780			0.1178			0.1440			0.0457		

* EQ means the equated score.

Table III.5: Equating Result for Reading Comprehension: 2016 ECZ vs. 2014

	Identity			Mean			Linear			Equipercntile			Circle-arc		
	2016 ECZ	2014	EQ*	2016 ECZ	2014	EQ*	2016 ECZ	2014	EQ*	2016 ECZ	2014	EQ*	2016 ECZ	2014	EQ*
Mean	1.971	1.794	1.971	1.971	1.794	1.794	1.971	1.794	1.794	1.971	1.794	1.784	1.971	1.794	1.856
SD	1.459	1.437	1.459	1.459	1.437	1.459	1.459	1.437	1.437	1.459	1.437	1.428	1.459	1.437	1.448
Skew	0.524	0.704	0.524	0.524	0.704	0.524	0.524	0.704	0.524	0.524	0.704	0.698	0.524	0.704	0.689
Kurt	2.530	2.827	2.530	2.530	2.827	2.530	2.530	2.827	2.530	2.530	2.827	2.879	2.530	2.827	2.753
Min	0.000	0.000	0.000	0.000	0.000	-0.177	0.000	0.000	-0.146	0.000	0.000	-0.075	0.000	0.000	0.000
Max	5.000	5.000	5.000	5.000	5.000	4.823	5.000	5.000	4.776	5.000	5.000	4.946	5.000	5.000	5.000
N	339	339	339	339	339	339	339	339	339	339	339	339	339	339	339
Standard Error of Equating	0.000			0.1024			0.1524			0.1683			0.0579		

* EQ means the equated score.

PART V. SUMMARY RESULTS

Table V.2: Mean Ratios for the Reading Comprehension Subtask

	Mean Base From	Mean New From	Mean Ratio
2016 TTL onto 2012	0.643	1.354	0.475
2016 ECZ onto 2014	1.794	1.971	0.910

Figure I.I: Histogram for Age

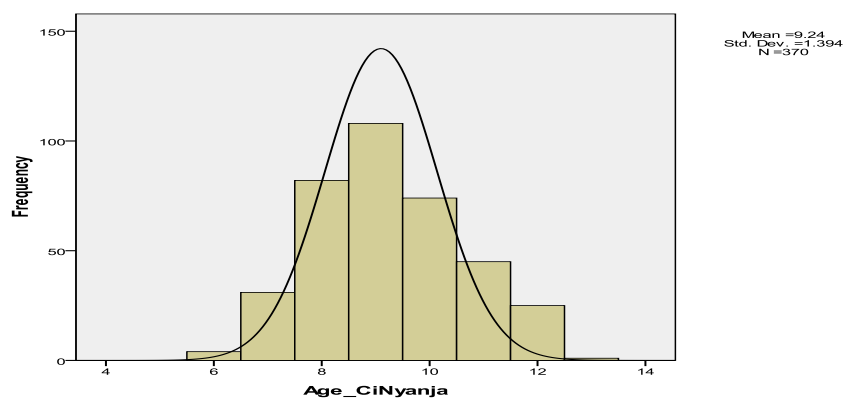


Figure II.I: Distributions for the Oral Reading Fluency Task

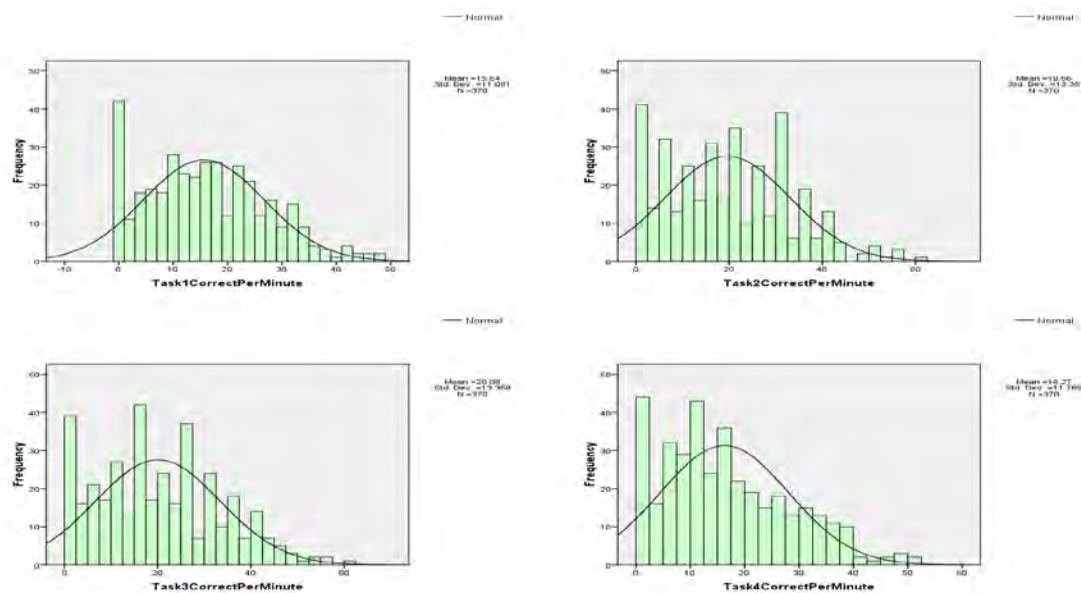


Figure II.2: Distributions for the Reading Comprehension Task

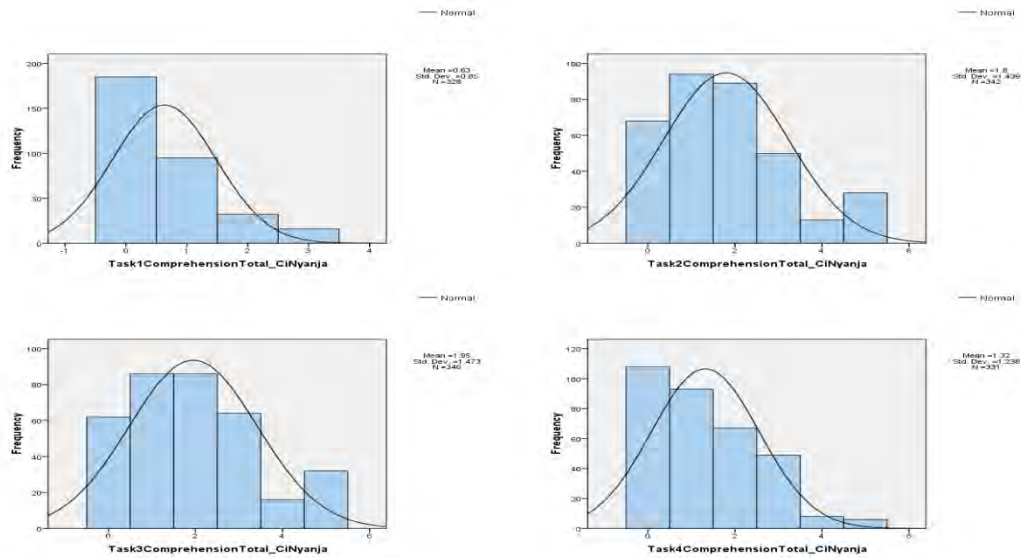


Figure II.3: Scatter Plot of Oral Reading Fluency: 2016 TTL vs. 2012 and 2016 ECZ vs. 2014

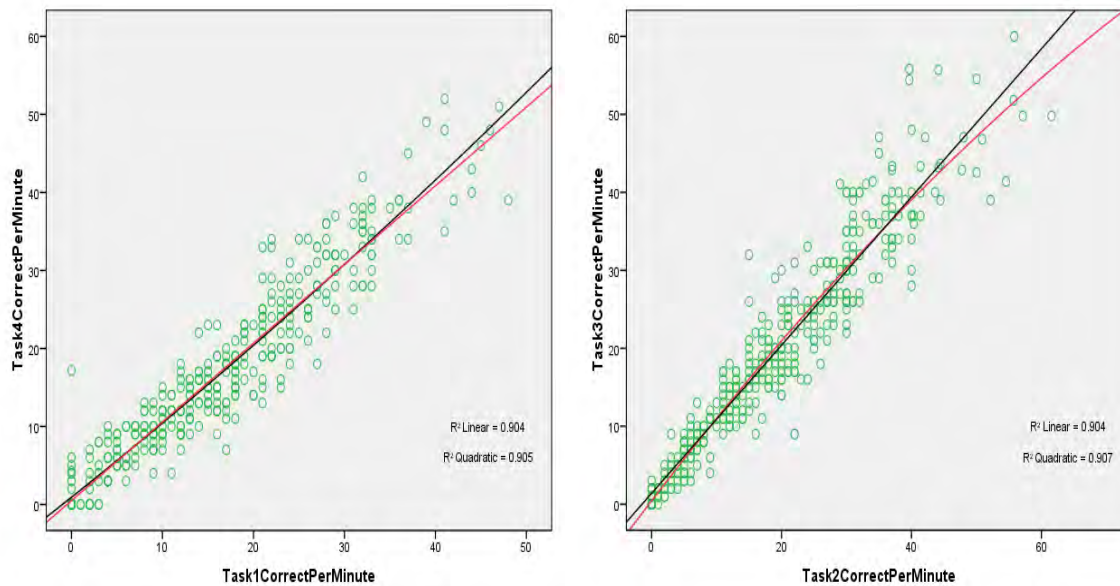
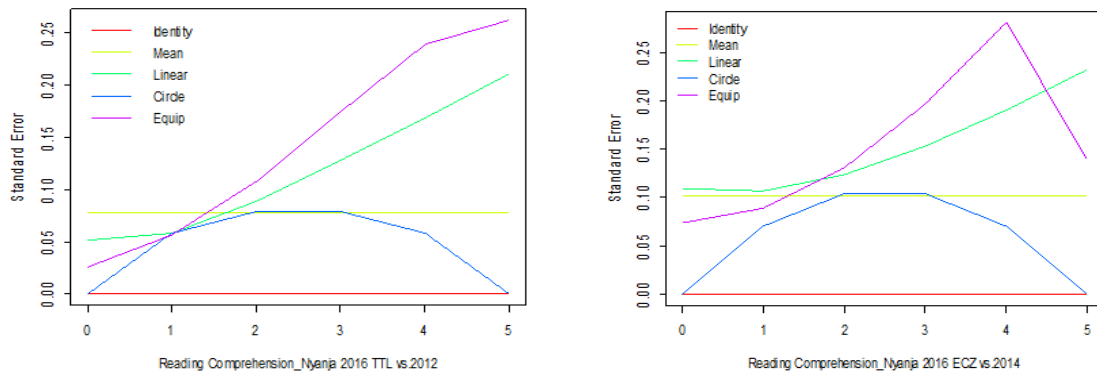


Figure III.3: Standard Error of Equating for Reading Comprehension: 2016 TTL vs. 2012 and 2016 ECZ vs. 2014



DETAILED RESULTS BY LANGUAGE: CHITONGA

PART I. SUMMARY OF DEMOGRAPHIC VARIABLES

Table I.1: Frequency for Sex – ChiTonga

Sex	Female	Male	Total
N	242	181	423

PART II. SUMMARY STATISTICS FOR THE TWO READING TASKS

Table II.1: Summary Statistics for the Oral Reading Fluency Task

	Fluency 2012	Fluency 2014	Fluency 2016 ECZ	Fluency 2016 TTL
N Valid*	423	423	422	423
Missing	0	0	1	0
Mean	18.71	20.17	20.58	20.98
Standard Deviation	11.670	12.641	12.007	11.858
Minimum	0	0	0	0
Maximum	59	68	60	59

* Indicates the total observations with seconds elapsed greater than 5.

Table II.2: Summary Statistics for the Reading Comprehension Task

	Comprehension 2012	Comprehension 2014	Comprehension 2016 ECZ	Comprehension 2016 TTL
N Valid*	384	392	389	395
Missing	39	31	34	28
Mean	1.19	1.48	1.33	1.44
Standard Deviation	1.134	1.338	1.295	1.044
Minimum	0	0	0	0
Maximum	5	5	5	5

* Indicates the total observations with seconds elapsed greater than 5.

Table II.3: Correlation Coefficients for the Two Reading Tasks

	Oral Reading Fluency		Reading Comprehension	
	2016 TTL vs. 2012	2016 ECZ vs. 2014	2016 TTL vs. 2012	2016 ECZ vs. 2014
Correlation*	0.930*	0.945*	0.647*	0.744*
Coefficient	(n=370)	(n=370)	(n=322)	(n=339)

* Correlation is significant at the 0.01 level (2-tailed). N is counted as non-missing scores for both tests.

PART III. EQUATE RESULTS FOR THE TWO READING TASKS

Table III.1: Mean Difference and Standard Error for the Oral Reading Fluency Task

	2012	2014	2016 ECZ	2016 TTL	Mean Difference	
					2016 TTL vs. 2012	2016 ECZ vs. 2014
Mean	18.71	20.17	20.58	20.98	2.27	0.41
Standard Deviation	11.67	12.641	12.007	11.858		
Standard Error	1.167	1.2641	1.2007	1.1858		

Table III.2: Mean Difference and Standard Error for the Reading Comprehension Task

	2012	2014	2016 ECZ	2016 TTL	Mean Difference	
					2016 TTL vs. 2012	2016 ECZ vs. 2014
Mean	1.19	1.48	1.33	1.44	0.25	-0.15
Standard Deviation	1.134	1.338	1.295	1.044		
Standard Error	0.1134	0.1338	0.1295	0.1044		

Table III.3: Equating Result for Oral Reading Fluency: 2016 TTL vs. 2012

	Identity			Mean			Linear			Equipercntile			Circle-arc		
	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*
Mean	23.249	20.87	23.249	23.249	20.87	20.87	23.249	20.87	20.87	23.249	20.87	20.844	23.249	20.87	21.174
SD	10.341	10.375	10.341	10.341	10.375	10.341	10.341	10.375	10.375	10.341	10.375	10.314	10.341	10.375	10.047
Skew	0.217	0.61	0.217	0.217	0.61	0.217	0.217	0.61	0.217	0.217	0.61	0.629	0.217	0.61	0.397
Kurt	2.966	3.282	2.966	2.966	3.282	2.966	2.966	3.282	2.966	2.966	3.282	3.77	2.966	3.282	2.966
Min	2	1	2	2	1	-0.378	2	1	-0.448	2	1	1.702	2	1	1.672
Max	59	59	59	59	59	56.622	59	59	56.739	59	59	59.37	59	59	59
N	378	378	378	378	378	378	378	378	378	378	378	378	378	378	378
Standard Error of Equating	0.000			0.7729			1.1644			1.5320			0.4548		

* EQ means the equated scores.

Table III.4: Equating Result for Reading Comprehension: 2016 TTL vs. 2012

	Identity			Mean			Linear			Equipercntile			Circle-arc		
	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*
Mean	1.504	1.203	1.504	1.504	1.203	1.203	1.504	1.203	1.203	1.504	1.203	1.14	1.504	1.203	1.263
SD	1.022	1.133	1.022	1.022	1.133	1.022	1.022	1.133	1.133	1.022	1.133	1.076	1.022	1.133	0.922
Skew	0.212	1.219	0.212	0.212	1.219	0.212	0.212	1.219	0.212	0.212	1.219	1.451	0.212	1.219	0.626
Kurt	3.03	4.171	3.03	3.03	4.171	3.03	3.03	4.171	3.03	3.03	4.171	6.724	3.03	4.171	3.03
Min	0	0	0	0	0	-0.301	0	0	-0.464	0	0	-0.179	0	0	0
Max	5	5	5	5	5	4.699	5	5	5.08	5	5	5.484	5	5	5
N	379	379	379	379	379	379	379	379	379	379	379	379	379	379	379
Standard Error of Equating	0.000			0.0812			0.1590			0.1344			0.0466		

* EQ means the equated scores.

Table III.5: Equating Result for Reading Comprehension: 2016 ECZ vs. 2014

	Identity			Mean			Linear			Equipercntile			Circle-arc		
	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*
Mean	1.343	1.499	1.343	1.343	1.499	1.499	1.343	1.499	1.499	1.343	1.499	1.492	1.343	1.499	1.446
SD	1.296	1.337	1.296	1.296	1.337	1.296	1.296	1.337	1.337	1.296	1.337	1.327	1.296	1.337	1.344
Skew	0.916	0.834	0.916	0.916	0.834	0.916	0.916	0.834	0.916	0.916	0.834	0.869	0.916	0.834	0.734
Kurt	3.411	3.311	3.411	3.411	3.311	3.411	3.411	3.311	3.411	3.411	3.311	3.263	3.411	3.311	3.411
Min	0	0	0	0	0	0.156	0	0	0.113	0	0	0.104	0	0	0
Max	5	5	5	5	5	5.156	5	5	5.272	5	5	5.141	5	5	5
N	385	385	385	385	385	385	385	385	385	385	385	385	385	385	385
Standard Error of Equating	0.000			0.0873			0.1602			0.1463			0.0694		

* EQ means the equated scores.

PART V. SUMMARY RESULTS

Table V.1: Mean Ratios for the Oral Reading Fluency Task

	Mean Base From	Mean New From	Mean Ratio
2016 TTL onto 2012	20.870	23.249	0.898

Table V.2: Mean Ratios for the Reading Comprehension Task

	Mean Base From	Mean New From	Mean Ratio
2016 TTL onto 2012	1.203	1.504	0.800
2016 ECZ onto 2014	1.499	1.343	1.116

Figure I.1: Histogram for Age

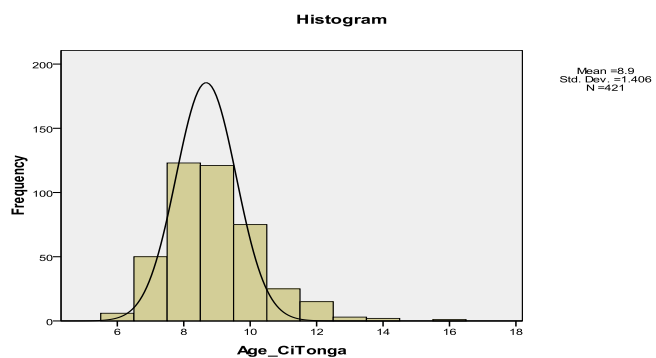


Figure II.1: Distributions for the Oral Reading Fluency Task

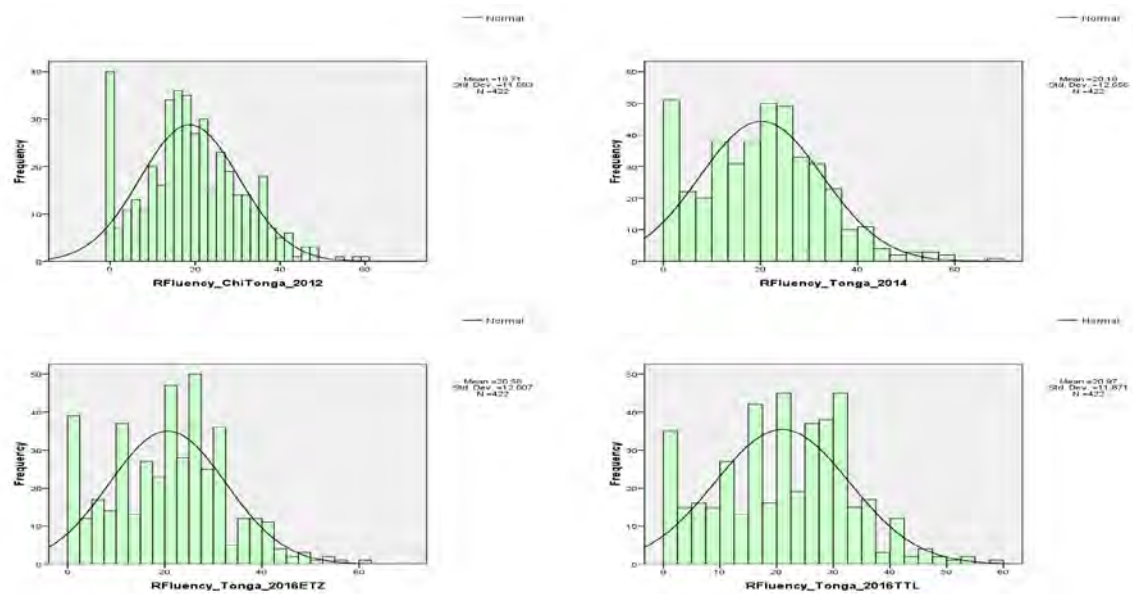


Figure II.2: Distributions for the Reading Comprehension Task

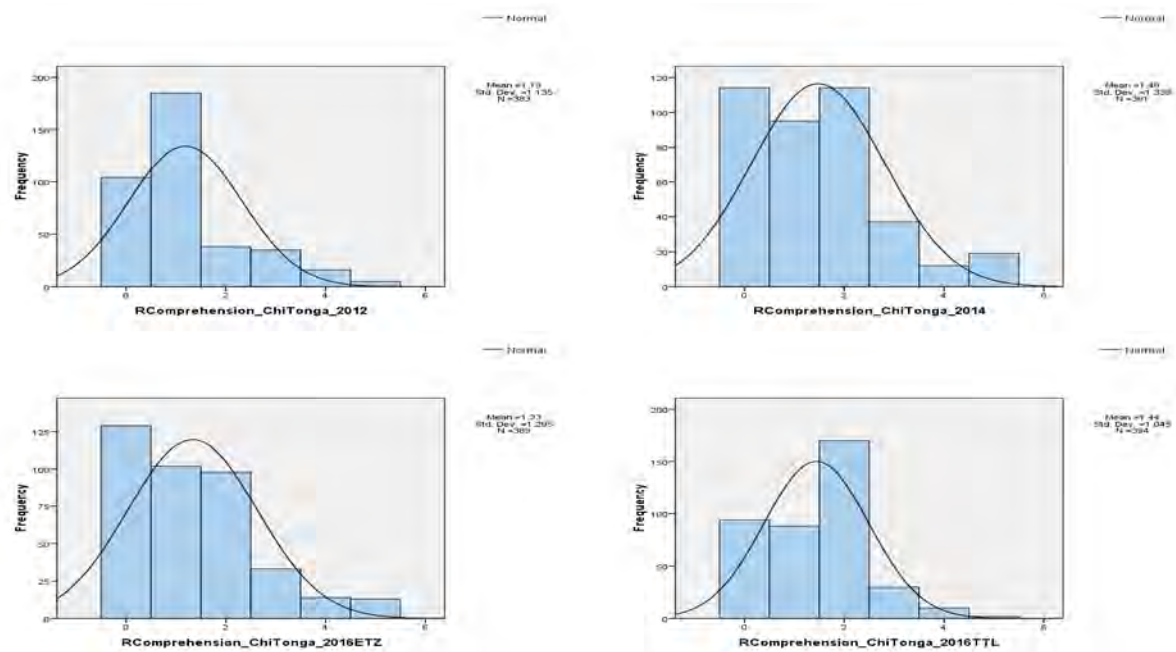


Figure II.3: Scatter Plot of Oral Reading Fluency: 2016 TTL vs. 2012 and 2016 ECZ vs. 2014

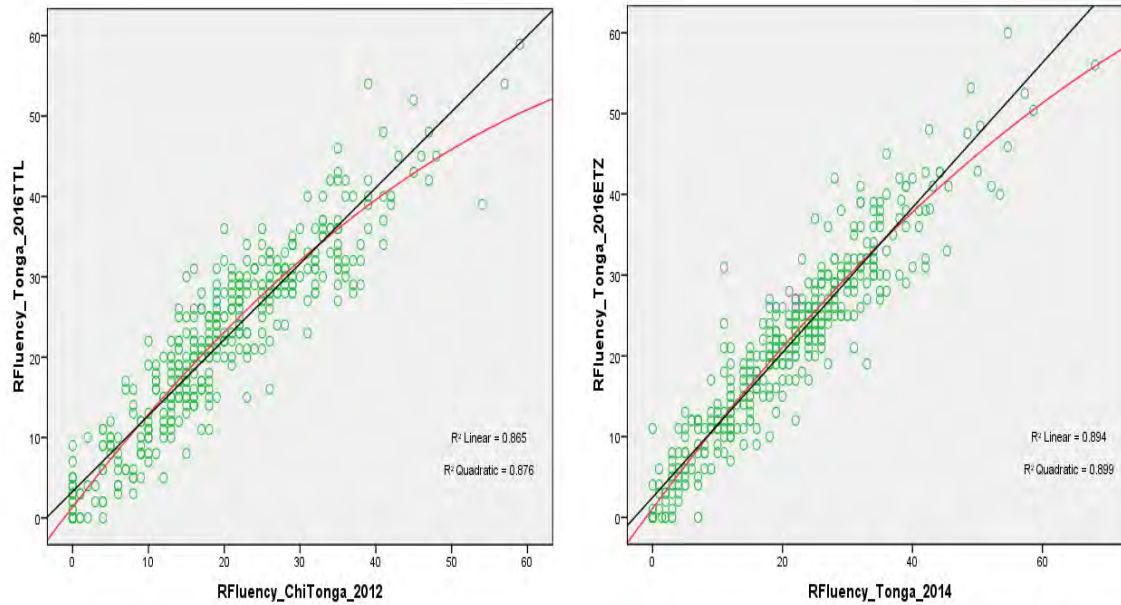


Figure III.1: Cumulative Distribution Plot of Oral Reading Fluency: 2016 TTL vs. 2012

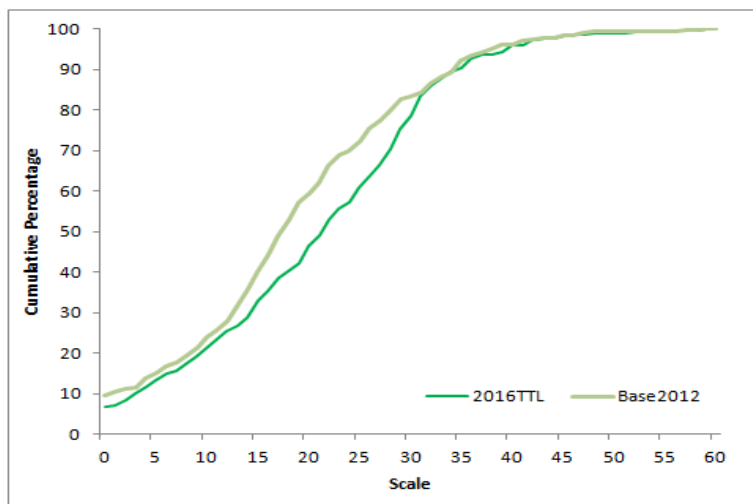


Figure III.2: Standard Error of Equating for Oral Reading Fluency: 2016 TTL vs. 2012

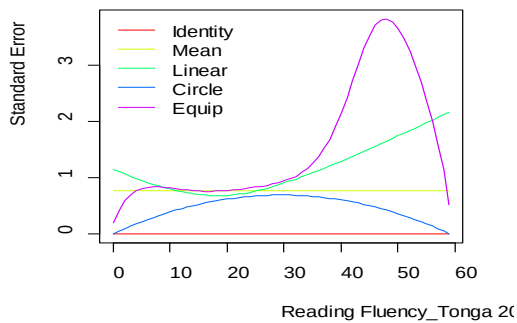
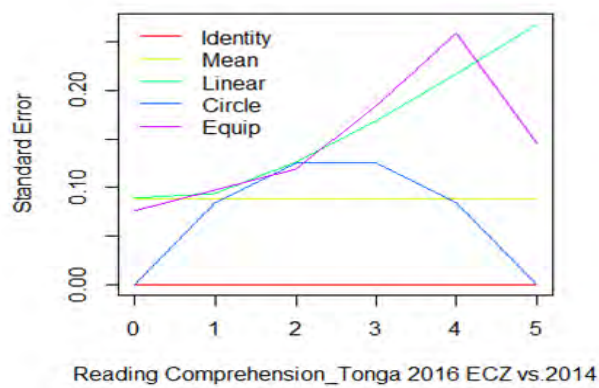
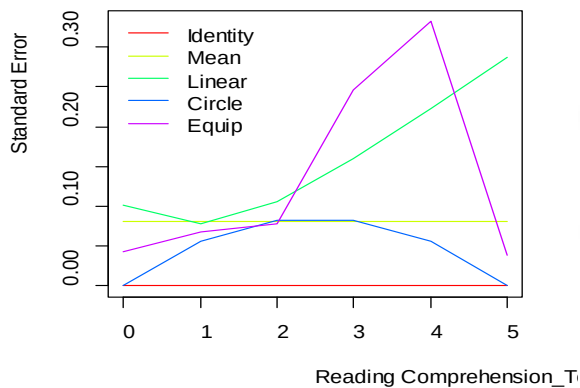


Figure III.3: Standard Error of Equating for Reading Comprehension: 2016 TTL vs. 2012 and 2016 ECZ vs. 2014



DETAILED RESULTS BY LANGUAGE: ICIBEMBA

PART I. SUMMARY OF DEMOGRAPHIC VARIABLES

Table I.2: Frequency for Sex – iCiBemba

Sex	Female	Male	Total
N	169	198	367

PART II. SUMMARY STATISTICS FOR THE TWO READING TASKS

Table II.1: Summary Statistics for the Oral Reading Fluency Task

	Fluency 2012	Fluency 2014	Fluency 2016 ECZ	Fluency 2016 TTL
N Valid*	361	361	361	361
Missing	0	0	1	0
Mean	19.13	23.52	22.96	17.55
Standard Deviation	12.334	14.117	13.822	11.948
Minimum	0	0	0	0
Maximum	58	71	86	58

* Indicates the total observations with seconds elapsed greater than 5.

Table II.2: Summary Statistics for the Reading Comprehension Task

	Comprehension 2012	Comprehension 2014	Comprehension 2016 ECZ	Comprehension 2016 TTL
N Valid*	316	326	326	318
Missing	51	41	41	49
Mean	1.19	1.41	1.30	1.47
Standard Deviation	1.038	1.353	1.368	1.065
Minimum	0	0	0	0
Maximum	3	4	5	5

* Indicates the total observations with seconds elapsed greater than 5.

Table II.3: Correlation Coefficients for the Oral Reading Fluency Task and Reading Comprehension Task

	Oral Reading Fluency		Reading Comprehension	
	2016 TTL vs. 2012	2016 ECZ vs. 2014	2016 TTL vs. 2012	2016 ECZ vs. 2014
Correlation	0.950*	0.949*	0.525*	0.7511*
Coefficient	(n=361)	(n=361)	(n=310)	(n=321)

* Correlation is significant at the 0.01 level (2-tailed). N is counted as non-missing scores for both tests.

PART III. EQUATE RESULTS FOR THE TWO READING TASKS

Table III.1: Mean Difference and Standard Error for the Oral Reading Fluency Task

	2012	2014	2016 ECZ	2016 TTL	Mean Difference	
					2016 TTL vs. 2012	2016 ECZ vs. 2014
Mean	19.13	23.52	22.96	17.55	-1.58	-0.56
Standard Deviation	12.334	14.117	13.822	11.948		
Standard Error	1.2334	1.1412	1.3822	1.1195		

Table III.2: Mean and Standard Error for the Reading Comprehension Task

	2012	2014	2016 ECZ	2016 TTL	Mean Difference	
					2016 TTL vs. 2012	2016 ECZ vs. 2014
Mean	1.19	1.41	1.3	1.47	0.28	-0.11
Standard Deviation	1.038	1.353	1.368	1.065		
Standard Error	0.1038	0.1353	0.1368	0.1065		

Table III.3: Equating Result for Oral Reading Fluency 2016 TTL vs. 2012

	Identity			Mean			Linear			Equipercentile			Circle-arc		
	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*
Mean	20.294	22.177	20.294	20.294	22.177	22.177	20.294	22.177	22.177	20.294	22.177	22.186	20.294	22.177	21.905
SD	10.552	10.514	10.552	10.552	10.514	10.552	10.552	10.514	10.514	10.552	10.514	10.487	10.552	10.514	10.861
Skew	0.543	0.479	0.543	0.543	0.479	0.543	0.543	0.479	0.543	0.543	0.479	0.492	0.543	0.479	0.404
Kurt	3.164	3.228	3.164	3.164	3.228	3.164	3.164	3.228	3.164	3.164	3.228	3.076	3.164	3.228	3.164
Min	1	1	1	1	1	2.884	1	1	2.954	1	1	1.954	1	1	1.141
Max	58	58	58	58	58	59.884	58	58	59.746	58	58	58.122	58	58	58
N	310	310	310	310	310	310	310	310	310	310	310	310	310	310	310
Standard Error of Equating	0.000			0.8223			1.3277			1.5820			0.5818		

* EQ means the equated scores.

Table III.4: Equating Result for Reading Comprehension 2016 TTL vs. 2012

	Identity			Mean			Linear			Equipercntile			Circle-arc		
	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*	2016 TTL	2012	EQ*
Mean	1.506	1.206	1.506	1.506	1.206	1.206	1.506	1.206	1.206	1.506	1.206	1.195	1.506	1.206	1.27
SD	1.054	1.032	1.054	1.054	1.032	1.054	1.054	1.032	1.032	1.054	1.032	1.011	1.054	1.032	0.954
Skew	0.239	0.375	0.239	0.239	0.375	0.239	0.239	0.375	0.239	0.239	0.375	0.276	0.239	0.375	0.567
Kurt	2.595	1.972	2.595	2.595	1.972	2.595	2.595	1.972	2.595	2.595	1.972	2.365	2.595	1.972	2.595
Min	0	0	0	0	0	-0.3	0	0	-0.267	0	0	-0.189	0	0	0
Max	5	3	5	5	3	4.7	5	3	4.624	5	3	4.444	5	3	5
N	310	310	310	310	310	310	310	310	310	310	310	310	310	310	310
Standard Error of Equating	0.000			0.0871			0.1217			0.1470			0.060		

* EQ means the equated scores.

PART V. SUMMARY RESULTS

Table V.1: Mean Ratios for the Oral Reading Fluency Task

	Mean Base From	Mean New From	Mean Ratio
2016 TTL onto 2012	22.177	20.294	1.093

Table V.2: Mean Ratios for the Reading Comprehension Task

	Mean Base From	Mean New From	Mean Ratio
2016 TTL onto 2012	1.206	1.506	0.801

Figure I.1: Histogram for Age

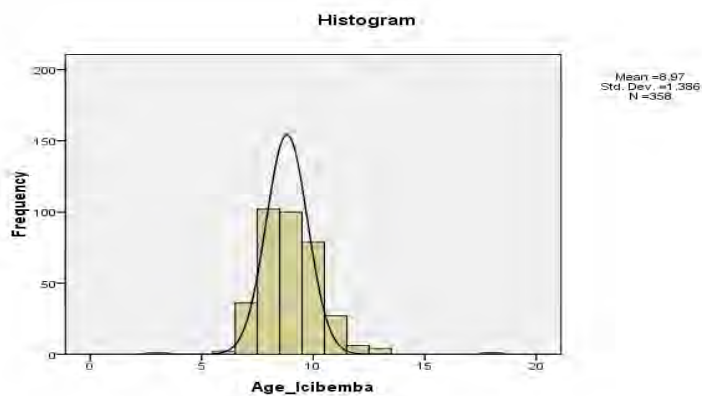


Figure II.1: Distributions for the Oral Reading Fluency Subtask

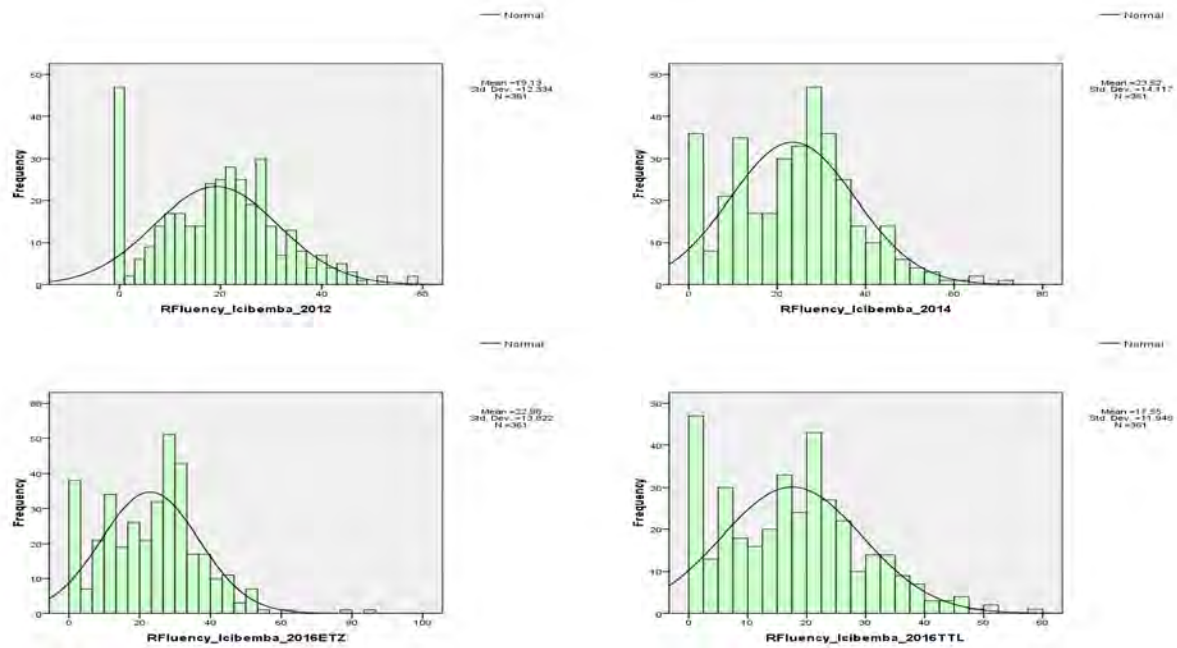


Figure II.2: Distributions for the Reading Comprehension Subtask

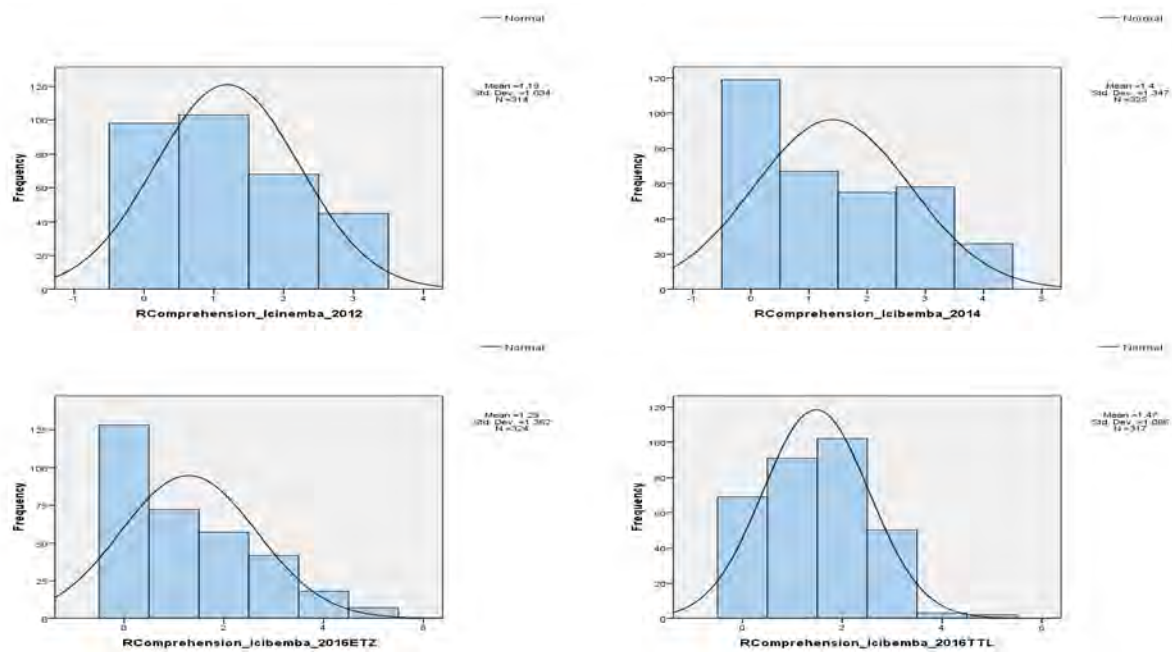


Figure II.3: Scatter Plot of Oral Reading Fluency: 2016 TTL vs. 2012 and 2016 ECZ vs. 2014

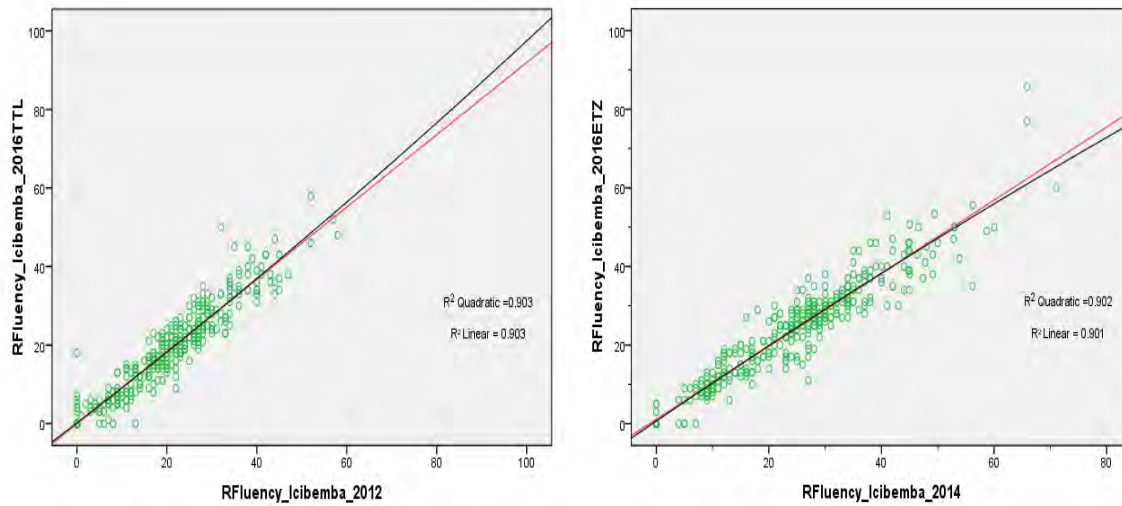


Figure III.1: Cumulative Distribution Plot of Oral Reading Fluency: 2016 TTL vs. 2012

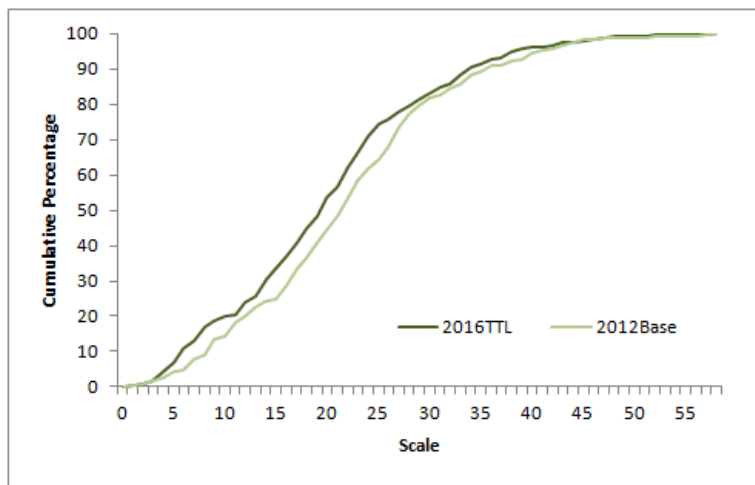


Figure III.2: Standard Error of Equating for Oral Reading Fluency: 2016 TTL vs. 2012

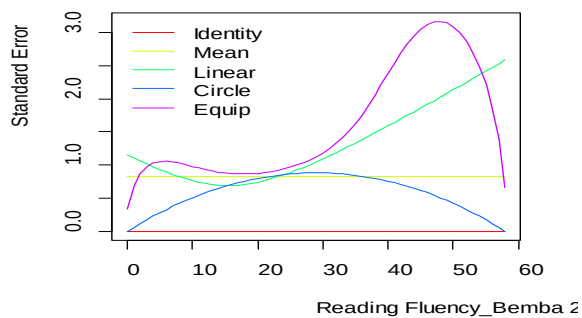
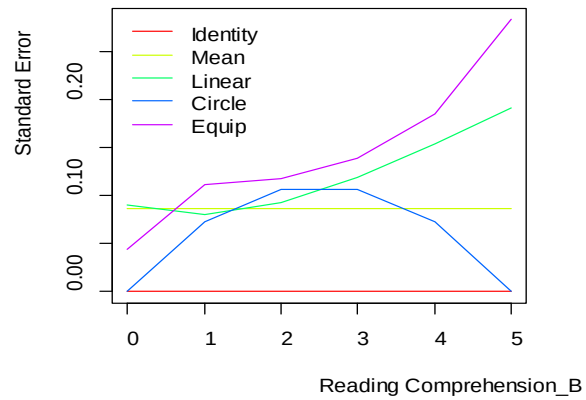


Figure III.3: Standard Error of Equating for Reading Comprehension: 2016 TTL vs. 2012 and 2016 ECZ vs. 2014



ANNEX 5. EVALUATION QUESTIONS BY DATA SOURCE AND INDICATOR

Key Evaluation Question	Sub-questions	Data Source	Indicator	Data Analysis
1. To what extent have TTL interventions improved literacy achievement in community schools? (TTL IR 2)	a. To what extent have TTL interventions increased reading skills? b. What proportion of learners in TTL-supported community schools can read and understand the meaning of grade-level text after 2 years of primary schooling?	EGRA	Number of male and female learners in grade 2 who exhibit reading skills gains on EGRA tasks by province	Descriptive statistics for each EGRA task by sex, province (and language ²²), with comparison to baseline and midline scores: - Mean percentage correct, noting significance of changes across phases (t-tests). Comparison of proportions across each evaluation phase for each EGRA task, noting differences in distributions: - Calculation of proportion showing reading skill gains using the MOGE benchmarks (tasks 2, 3, 5a, 5b).

²² Note that the sample is not drawn to be representative by language, because not all schools using that language of instruction are in the sample frame; region is the main driver of the sampling frame. See the explanation in 3.1, "Sample."

Key Evaluation Question	Sub-questions	Data Source	Indicator	Data Analysis
2. How has the MOGE's engagement with community schools changed as a result of TTL interventions? (TTL IR 1)	a. To what extent are community schools receiving more support from the MOGE?	Community School Head Teacher Questionnaire	Forms of support: ²³ 1. Infrastructure 2. Grants ²⁴ 3. Teaching and learning materials	Descriptive statistics of each form of support at the six-province aggregate level, compared to baseline and midline (reported separately by head teachers and District Education Board Secretaries): - Median number of forms of support provided per school (significance of changes calculated using Mann-Whitney U) - Proportion of schools receiving each form of support - Proportion of schools receiving at least one form of support. Descriptive statistics of MOGE monitoring activities (reported by head teachers and teachers).
	b. To what extent has the MOGE improved its monitoring of community schools?	District Education Board Secretary Questionnaire Classroom Observation Protocol Interview	4. Free basic materials (chalk, notebooks, pens/pencils, etc.) 5. CPD 6. Government-seconded teachers 7. Monitoring	
3. To what extent are teachers implementing TTL-supported literacy teaching methods? (TTL IR 2)		Classroom Observation Protocol Community School Head Teacher Questionnaire		Comparison to midline using significance tests for each criterion for the following indicators: - Mean number of intervals during which each criterion was performed - Percent of lessons fulfilling the criterion at least once Descriptive statistics of instructional and administrative practice (reported by head teachers and teachers)

²³ Indicators 1–6 correspond to evaluation question 2.a, and number 7 informs evaluation question 2.b.

²⁴ Allocated through the District Education Board Secretary, grants to community schools are earmarked for infrastructure maintenance, administration, teaching and learning materials, support to orphans and vulnerable children, and school health and nutrition (MOGE 2016b).

ANNEX 6. REFERENCES

- Angoff, W. H. 1984. *Scales, Norms, and Equivalent Scores*. Princeton, NJ: Educational Testing Service.
- Beatty, A. and L. Pritchett. 2012. *From Schooling Goals to Learning Goals: How Fast Can Student Learning Improve?* CGD Policy Paper 012. Washington, D.C.: Center for Global Development. Retrieved from: <http://www.cgdev.org/content/publications/detail/1426531> (updated January 29, 2013).
- Falconer-Stout, Z., E. Eunifidah, and T. Mayapi. 2014. *Government Teachers in Community Schools: Two Zambian Success Stories*. Time to Learn Project. Lusaka, Zambia: USAID. Retrieved from: https://encompassworld.com/sites/default/files/working_relationships_final_public_usaid_approved.pdf
- Falconer-Stout, Z. J. and K. Kalimaposo. 2014. *Active Parents, Active Learners? Lessons from Two Community Schools in Zambia*. Time to Learn Case Study Series. Time to Learn Project. Lusaka, Zambia: USAID. Retrieved from: https://encompassworld.com/sites/default/files/active_pcscs_final_public_usaid_approved_0.pdf
- Frischkorn, R., and Z. Falconer-Stout. 2016. *Ensuring Inclusive and Quality Education for All: A Comprehensive Review of Community Schools in Zambia*. Time to Learn Project. Lusaka, Zambia: USAID. Retrieved from: https://encompassworld.com/sites/default/files/Review-Community-Schools-Zambia-FINAL_PUBLIC-%2809.01.2016%29_0.pdf
- Frischkorn, R., Z. Falconer-Stout, and L. Messner. 2016. *Time to Learn Project: Year 4 Performance Evaluation Report*. Time to Learn Project. Lusaka, Zambia: USAID. Retrieved from: https://encompassworld.com/sites/default/files/TTL_PY4_Perf-Eval-Report_FINAL_3March2016.pdf
- Hanson, B. A., L. Zeng, and D. A. Colton. 1994. *A Comparison of Presmoothing and Postsmoothing Methods in Equipercentile Equating*. Iowa City, Iowa: American College Testing Program.
- Holland, P. W., and N. J. Dorans. 2006. Linking and Equating. In R. L. Brennan, ed. *Educational measurement* (4th ed.). Westport, CT: Greenwood, 187–220.
- Kim, Y.-S. G., H. N. Boyle, S. S. Zuilkowski, and P. Nakamura. 2016. *Landscape Report on Early Grade Literacy*. Washington, D.C.: USAID. Retrieved from: https://globalreadingnetwork.net/sites/default/files/research_files/LandscapeReport.pdf
- Kolen, M. J. 1988. Traditional equating methodology. *Educational Measurement: Issues and Practice* 7(4): 29–37.
- Kolen, M. J., and R. L. Brennan. 2004. *Test equating, scaling, and linking: Method and practice* (2nd ed.). New York: Springer-Verlag.
- Livingston, S. A. 2004. *Equating Test Scores (without IRT)*. Princeton, NJ: Educational Testing Service.
- Livingston, S. A., and S. Kim. 2009. The circle-arc method for equating in small samples. *Journal of Educational Measurement* 46(3): 330–343.
- Livingston, S. A. and S. Kim. 2010. Random-groups equating with samples of 50 to 400 test takers. *Journal of Educational Measurement* 47:175–185
- Ministry of Education, Science, Vocational Training, and Early Education. 2014. *Reading Performance Level Descriptors for Grades 1–4*. Lusaka, Zambia: MOGE.
- MOGE. 2016a. 2015. *Educational Statistical Bulletin*. Lusaka, Zambia: MOGE.

- MOGE. 2016b. *Operational Guidelines of Community Schools—2014*. Lusaka, Zambia: MOGE.
- MOGE. 2015a. *Proposing Benchmarks and Targets for Early Grade Reading and Mathematics in Zambia*. Lusaka, Zambia: MOGE.
- MOGE. 2015b. *Zambia's National Assessment Survey Report 2014: Learning Achievement at the Primary School Level*. Lusaka, Zambia: MOGE.
- MOGE 2013a. *National Literacy Framework*. Lusaka, Zambia: MOGE.
- MOGE 2013b. *Zambian Languages Syllabus: Grade 1–7*. Lusaka, Zambia: MOGE.
- Musonda, B., and A. Kaba. n.d. *The SACMEQ-III Project in Zambia: A Study of the Conditions of Schooling and the Quality of Education*. Southern and Eastern African Consortium for Monitoring Educational Quality Testing Consortium.
- Piper, B., S. S. Zuilkowski, and S. Ong'ele. 2016. Implementing Mother Tongue Instruction in the Real World: Results from a Medium-Scale Randomized Controlled Trial in Kenya. *Comparative Education Review* 60(4): 776–807.
- RTI International. 2015. *Early Grade Reading Assessment Toolkit, Second Edition*. Washington, D.C: USAID. Retrieved from: https://globalreadingnetwork.net/sites/default/files/resource_files/EGRA%20Toolkit%20Second%20Edition.pdf
- Skaggs, G. 2005. Accuracy of Random Groups Equating with Very Small Samples. *Journal of Educational Measurement* 42(4): 309–330.

ANNEX 7. DATA COLLECTION TOOLS

The complete set of tools used for data collection are provided under separate cover in conjunction with this report. Please refer to the supplemental document, “Time to Learn Endline Evaluation Report: Annex 7,” for these tools.

TIME TO LEARN PROJECT
PLOT NUMBER 203B, OFF KUDU ROAD
KABULONGA, LUSAKA
ZAMBIA